DOCUMENT RESUME

ED 424 768 FL 025 559

AUTHOR	Tsagari, Constance
TITLE	Method Effect on Testing Reading Comprehension: How Far Can We Go?
PUB DATE	1994-08-00
NOTE	96p.; Master's Thesis, University of Lancaster.
PUB TYPE	Dissertations/Theses - Masters Theses (042)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Comparative Analysis; *English (Second Language); Foreign Countries; Language Processing; Language Tests; Multiple Choice Tests; Protocol Analysis; Questionnaires; *Reading Comprehension; *Reading Strategies; *Reading Tests; Second Language Instruction; Student Attitudes; *Test Format; *Test Wiseness

ABSTRACT

This study investigated and compared the effects of two test formats (free response and multiple choice) on English-as-a-Second-Language (ESL) learners' reading comprehension. The tests, together with a checklist of test-taking strategies and retrospective questionnaires concerning more general reading strategies, were administered to 57 ESL graduate students. Two students were also given think-aloud interviews on the tests. Results indicate that the two tests, with identical content but different formats, may not yield measures of the same trait. This was further evidenced by the frequency with which students selected strategies from the checklist to describe their ways of processing the same items in the two formats. A double-check on the checklist's validity showed it to be reflective of the students' strategies. Implications for test designers, teachers creating their own tests, and test validators are discussed. Appended are the test instruments and results. (Contains 41 references.) (Author/MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 424 768

UNIVERSITY OF LANCASTER

DEPARTMENT OF LINGUISTICS AND MODERN ENGLISH LANGUAGE

METHOD EFFECT ON TESTING READING COMPREHENSION: HOW FAR CAN WE GO?

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Constance Tsagari

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CONSTANCE TSAGARI

M. A. in LINGUISTICS for ELT

August 1994

FL025559

BEST COPY AVAILABLE

Contents

CONTENTS i

ACKNOWLEDGEMENTS iii

ABSTRACT iv

CHAPTER 1 : INTRODUCTION 1

CHAPTER 2 : A REVIEW OF THE LITERATURE 3

2. 1 Inherent Aspects of the Two Methods Used 3

2. 2 Review of Research Findings on Comparison of Testing Methods 4

2. 3 Research Findings on Comparison of Free-response and Multiple-choice Formats 5

2. 4 Research Findings Utilising Verbal Reports as Data 8

2. 5 The Use of Self-report Instruments as a Means of Collecting Process Data 11

CHAPTER 3 : PURPOSES AND DESIGN OF THE PRESENT STUDY 15

3. 1 Purposes of the Study 15

3. 2 Research Design 16

3. 2. 1 The Materials 16

3. 2. 2 Designing the Checklist of Test-taking Strategies 17

3. 2. 3 Piloting the Instruments 17

3. 2. 4 The Students 18

3. 2. 5 The Instruments: Rational and Administration 18

3. 2. 6 The Scoring Procedure 22

3. 2. 7 The Introspective Interviews 22

3. 2. 7. 1 The Purpose 22

3. 2. 7. 2 The Subjects, Materials and Administration 22

3. 2. 7. 3 The Training Sessions 23

3. 2. 7. 4 The Procedure followed during the Interviews 23

3. 2. 7. 5 The Analysis of the Verbal Protocols 24

CHAPTER 4 : ANALYSIS OF THE RESULTS 25

4. 1 Quantitative Analysis of the Results 25

4. 2 Qualitative Analysis of the Results 28

4. 2. 1 Strategy Use per Question 30

4. 2. 2 Overall Frequency of Strategies 38

4. 2. 3 A General Comment on the Validity of the Checklist 41

4. 3 General Reading Strategies and Students' Perceptions of the Two Formats 43

CHAPTER 5 : GENERAL DISCUSSION OF THE RESULTS 45

CHAPTER 6 : CONCLUSION 50

BIBLIOGRAPHY 52

APPENDIX A 55

APPENDIX B 63

APPENDIX C 70

APPENDIX D 74

APPENDIX E 78

APPENDIX F 84

Acknowledgements

I wish to thank my supervisor, Prof. J. Charles Alderson, for introducing me to the world of testing and for his guidance and assistance with this study.

I would also like to thank Prof. C. Evangelides for his help during the collection of the data as well as A. Waters at the Institute for English Language Education for his support.

Special thanks also to the five M.A. students at the Linguistics Department and the two graduate students from the English Department at the University of Athens who volunteered to take part in the interviews. Thanks should also go to the rest 57 graduate students, without whose participation this study would not have been completed.

Finally, I would like to express my gratitude to all those who have been "friends in need", my colleagues and the staff from the Linguistics Department. My gratitude to my family and friends in Greece who have always been so supportive and understanding.

Constance Tsagari

M. A. in Linguistics for ELT

August 1994

METHOD EFFECT ON TESTING READING COMPREHENSION : HOW FAR CAN WE GO?

Abstract

The purpose of this study was to explore and examine the nature of test method effect on reading comprehension by using both a product- and a process-based approach. The two methods compared were free-response and multiple-choice format. The tests, together with a Checklist of test-taking strategies and Retrospective Questionnaires relating to more general reading strategies, task difficulty and students' perceptions of the two formats were administered to 57 non-native speakers of English. They were asked to take the two tests and introspect on their test-taking strategies immediately after the completion of every question. At the end of every test, they were also asked to fill in a Retrospective Questionnaire. Two of those students took part in think-aloud interviews on the same tests to measure the validity of the self-report instrument.

The analysis of the scores on the two tests showed that two tests with identical content but different formats may not yield measures of the same trait. This was further evidenced by the frequency with which students selected strategies from the Checklist in order to describe their ways of processing the same items in the two formats. The double-check on the validity of the Checklist of test-taking strategies proved that the instrument was quite reflective of the students' strategies.

These findings have major implications for test designers and teachers who devise their own tests and test validators who employ this kind of instruments for the collection of process data.

CHAPTER 1

INTRODUCTION

The study of English as a Foreign Language (EFL) has become almost widespread and forms part of the curriculum of many educational systems. During the course of their studies, language students are subjected to a variety of local and international tests aiming at evaluating their language proficiency. Many of those tests incorporate among other language elements and skills, reading comprehension papers which purport to measure candidates' ability to read and comprehend in EFL. Test results are often used as the basis for decision making and selection in a variety of academic and professional contexts. For this reason it is important that reading comprehension tests upon which such decisions are made, render an accurate measurement of the candidate's competence in that specific skill.

To reach, though, an accurate or at least adequately workable definition of reading comprehension, although said at the beginning of this century:

> " ... would almost be the acme of a psychologist's dream for it would be to ... unravel the tangled story of the most remarkable specific performance that civilisation has learned in all its history " (Huey 1908, quoted in Allan, 1992: 4)

Test designers, in their attempt to tap the skill(s) of reading comprehension, have developed a multitude of methods, the aim of which is to measure the extent to which a particular text has been understood.

The question is whether these different procedures are all testing the unobservable mental operations involved in reading comprehension to the same degree since they seem to be the vehicle through which information about it can be obtained. This places importance on the testing methods employed. If these interfere with the assessment of the construct of reading comprehension and affect the results obtained by them, invalid information is obtained, referred to as "method effect". Its extent has been difficult to measure through psychometric approaches only which would normally focus on the product rather than the process of arriving at it.

Because of the important decisions made on the basis of reading comprehension tests and the impact they have on the EFL student, it is important to investigate the quality of methods used to measure the students' ability to comprehend EFL texts.

The purpose of this study is to investigate the effect that two commonly used testing techniques have on the assessment of reading comprehension of an EFL text. More specifically, the methods employed are multiple-choice format and its equivalent counterpart, free-response item format which are both still widely employed by testers. The study is based both on a product-oriented but mainly on a process-based approach which by means of a self-report instrument hopes to come closer to the test-taking process of as big a number of test-takers as possible.

CHAPTER 2

A REVIEW OF THE LITERATURE

Before the discussion of the research findings in the field of test method effect on EFL reading comprehension tests, it is necessary to present a brief overview of the inherent properties of the two methods under examination.

2. 1 Inherent Aspects of the Two Methods Used.

Various testing methods are currently used to test reading comprehension in both large scale tests and in the classroom. The test methods selected for this study were multiple-choice and free-response questions for the diversity of opinions they have evoked and the frequency of use in reading comprehension tests.

Multiple-choice items consist of a question or statement posed in a stem followed by three, four or sometimes five alternative answers. Of these, only one is correct. Free-response items require an answer of a few sentences in response to a posed question.

Of interest is to look briefly at three aspects in an attempt to compare and contrast the demands each makes on both the students and their teachers or test-designers.

The employment of free-response method is seen as a process of constantly asking and answering questions inherent in the reading process and therefore parallel to an authentic task for testing this skill. Contrary to that, having to choose one out of four possible answers is rare if not a "confusing dilemma ... (which) ... runs contrary to the very idea of education" (Oller, 1979: 256).

It appears to be then that the most probable behaviour elicited by the multiple-choice format is a recognition, identification/discrimination and selection pattern. On the other hand, the candidate has to recall, search and use productive skills in order to provide an answer for free-response items. Thus a potential source of variance lies in these two reading tasks which seems to make different demands on the examinees' processing strategies.

From the point of view of reliability, multiple-choice items can be objectively scored even by a machine or a computer, yielding consistent scores within a restricted amount of time. Free-response reading items have lower reliability since the judgmental assessment of the answer provided by the examinees comes into play and their scoring could be expensive in terms of time and money.

Practically speaking, free-response questions are easier to construct while multiple-choice items require appropriate composition of distractors which need to be piloted first in terms of their effectiveness.

The choice of each of two seems to be made more in terms of its administration efficiency or intra- and inter-rater reliability rather than considering the test-taking processings involved in each and consequently the patterns of behaviour evoked in their use. The present study attempts to look into such patterns of behaviour and find out how different or similar these are.

2. 2 Review of Research Findings on Comparison of Testing Methods.

The question of whether item format affects the way examinees respond has been of considerable interest to test validators.

Bachman and Palmer (1981) employed a multi-trait multi-method matrix design to investigate the effect of three testing methods (oral interview, reading translation and self-rating) on assessment of the traits of reading and speaking skills. The results demonstrated that the scores were more influenced by the method of measurement than by the trait being assessed.

In a comparison of three methods used for testing listening comprehension, De Jong (1983, quoted in Gordon, 1987:19-20) reached the conclusion that True/False and modified cloze items provided the best assessment on listening comprehension than multiple-choice items. These were considered as having given too many clues to the testees.

Henning (1983), reported differences in validity indices among three oral testing methods, suggesting that all the procedures used might not be measuring the same

aspect of oral proficiency. In a similar attempt to test oral proficiency by the use of four methods, Shohamy, Reves and Bejarano (1986) have reported differences in scores obtained among the methods used.

Alderson and Urquhart (1988), testing four groups of non-native students using free-response and gap-filling formats on five passages which varied in subject area and difficulty, concluded that "there is evidence of a strong method effect" (ibid: 179).

All this research has demonstrated that different methods attempting to test the same trait may yield different scores which actually makes a strong claim for the existence of such a sensitive issue as the method effect of the testing techniques used. More specifically, in the case of the two methods under investigation, namely free-response and multiple-choice format, studies which show whether there is an influence of these two formats on the construct purported to measure, have generated contradicting results.

2. 3 Research Findings on Comparison of Free-response and Multiple-choice Formats.

Investigators, such as Patterson (1926) and Bracht and Hopkins (1970) have computed correlations between multiple-choice and free-response versions of a test and obtained high correlations (often in the order of .90). These correlations were generally interpreted to mean that format does not influence what a test measures. Other researchers have carried out more sophisticated analyses with mixed results.

Vernon (1962) designed a study aiming at distinguishing between content and format factors in vocabulary and reading tests. Parallel forms of several tests were developed in both multiple-choice and free-response formats and were administered to college students. The results showed no evidence of format factors and he concluded that questions expressed in multiple-choice and free-response formats did not measure different abilities.

However, French (1965) setting off to investigate if "test problems ... measure something different for examinees who solve them by using different methods" (ibid:10), noticed two contrasting problem-solving styles employed in the two formats:

> " ..... the use of some kind of reasoned or systematic approach as contrasted to less orderly scanning and visualising, with reliance on common sense " (ibid: 26).

In a later study, Traub and Fisher (1977) attempted to assess whether tests with identical content but different formats measure the same attribute. Among the formats employed were free-response and multiple-choice. One of the main conclusions that the researchers arrived at was that "different formats may yield measures of different abilities" (ibid: 367).

An investigation between the relationship of construct validity and test item format was that done by Ward, Frederiksen and Carlson (1980). Their test, was designed to throw light on the cognitive activities of taking the same test in both free-response and in multiple-choice form. After statistical analysis of the results, including factor analysis, the researchers concluded that :

> " The production of ideas depends heavily on abilities other than those which determine performance when the subject has only to evaluate alternatives which are presented for choice" (ibid: 27)

This investigation was extended later by Frederiksen, Ward, Case, Carlson and Samph (1981). The results of the analysis indicated that multiple-choice format does not measure the same cognitive skills measured by similar problems in free-response form. Since both investigations attempted to reach the cognitive processes of test-taking, it is perhaps frustrating that neither has included verbal report data from the test-takers themselves so as to triangulate the approach.

Contrary to that, in an individual investigation of three types of verbal items measured by different format, including multiple-choice and free-response format, Ward (1982) showed that format made no difference in the attribute measured.

Despite the contradictory results to previous findings, Ward's research demonstrated the need to compare a number of items in multiple-choice and free-

response formats to distinguish what an item format demands from an examinee in terms of solving a problem and in expressing its solution.

In another study, Samson (1983, quoted in Gordon, 1987: 20-21) compared multiple-choice, free-response and summary to determine which of these methods would provide a purer measure of reading comprehension. It was shown that although all three methods seemed to be measuring the construct of reading comprehension, the difficulty level of questions in the different methods varied, suggesting an effect of testing method on test scores.

In a similar research where three methods (multiple-choice, free-response and cloze) were used to test the reading comprehension of non-native speakers of English, Lewkowicz (1983) concluded that the correlations obtained between the different methods suggested that these "were measuring traits specific to the method in addition to a common trait or skills" (ibid: 47).

The study was, however, limited in that it did not look at qualitative evidence to further support the apparent method effect. Aware of that, the researcher further suggested the need for validational studies in tests of reading comprehension and also in other skills, in an attempt to contribute to our understanding of method and trait and consequently result in better tests.

Shohamy (1984) reports results of a study which examined the effect that multiple-choice and free-response methods have on the measurement of reading comprehension. The results obtained in this study point to differences in students' scores on reading comprehension as a result also of a different testing method. Shohamy concludes that before anything can be said for sure about what is actually involved in doing these two kind of tasks, research should be done into the process of test taking " in order to try and explain the processes which the test-taker goes through in doing multiple-choice and free-response testing tasks" (ibid: 157). She goes on recommending the procedures of introspection as a methodological tool.

Research has also suggested that differences in scores occur as a result of variations within the same method (Alderson, 1983; Klein-Braley, 1983; Shohamy, 1983;

Bachman, 1985, etc.) which, although not being the focus of this study, further supports the assumption that testing methods seem to have an effect on the information we receive from tests.

The correlational research cited above mainly shows that different methods attempting to test the same trait may yield different scores in some cases while it may not in others, thus giving inconclusive results. Researchers, up to the point discussed, have limited the validity of their findings by adopting only empirical data through statistical approaches, particularly in the case of factor analysis, in order to detect different patterns of responses in test methods. A possibly more profitable approach, as already suggested by some of the investigators above (Lewkowicz, 1983; Shohamy, 1984), would be to collect qualitative data, rather than quantitative only, of the test-taking process or of how students interact with the experience of taking free-response and multiple-choice format reading test items. A next step would be to compare these so as to establish a possibly more valid approach to the trait each method is testing. It may then be feasible to find out if different test methods require examinees to follow different mental paths in their attempt to answer reading test items.

2. 4 Research Findings Utilising Verbal Reports as Data.

Contrary to psychometric procedures, verbal report methodology was proposed as a way of gathering data about examinees' problem-solving strategies in order to establish construct validity of the trait being measured.

Psychologists in the late 19th and early 20th centuries used the method in order to come closer to the way the human mind works. Verbal reports in the form of introspections (operations during which a reader verbalises his/her thought processes while taking a given task) and retrospective reports (which probe the subject for information after the completion of the task-induced process) were severely criticised with the advent of the influential behaviourist school. They emerged again with the development of the cognitive science for investigating reading and problem-solving processes (Ericsson and Simon, 1980).

A growing body of research has moved away from focus on product to investigating the reading process in order to define the reading comprehension trait both in L1 and L2 (Olshavski, 1977; Cohen and Hosenfeld, 1981; Mann, 1982; Cavalcanti, 1982; Afflerbach and Johnston, 1984; Bereiter and Bird, 1985; Block, 1986; Cohen, 1986; Grotjahn, 1987; Sarig, 1987; Pritchard, 1990).

Only recently has language testing research turned to the examination of processes involved in the test-taking experience. Cohen (1984) issued a call for the inclusion of such data in an attempt to "explore the closeness-of-fit between the tester's *presumptions* about what is being tested and the *actual* processes that the test-taker goes through" (ibid: 70).

Dollerup, Glahn and Hansen (1982), used introspective data to investigate students' reading strategies and the test solving techniques they employed on "Sprogtest" - a set of multiple-choice questions embedded in a continuous reading comprehension text. Dollerup et al reported that the students were asked to take the test and give reasons for their choice and in some cases to expand on their explanations. The "explanations" (ibid: 94) were analysed and it was concluded that reading-test solving strategies and techniques do to some extent overlap with the reading process as such. Two interesting problems were observed during the analysis of the "explanations": the first was that few of the participants managed to elicit "clean" (ibid) answers, by which they meant that more than one reading strategy brought readers to the answer. The other problem was the observation that "erroneous decoding will sometimes lead to correct answer" (ibid).

Farr, Pritchard and Smitten (1990) also investigated, the reading and test-taking strategies college students used to complete a portion of a multiple-choice reading comprehension test. Special care was taken over the procedures of gathering process data and the analysis of the protocols. The discussion of the analysis revealed many interesting patterns about how examinees process multiple-choice reading items making the researchers wonder whether these specific findings could be generalised to other kinds of reading tasks, reading-age groups and readers with different abilities.

The research of Farr et al has again emphasised the need for researchers and theorists not to be limited to product information only in order to infer process behaviour but to collect on-line data.

Anderson, Bachman, Perkins and Cohen (1991) report the use of a variety of data - both think-aloud protocols and more commonly used types of information on test content and test performance - in order to investigate the construct validity of a multiple-choice reading comprehension test. The end product was a list of 47 "Processing strategies" that students, according to the researchers, employ in order to give answers to multiple-choice tests. Its number and division into five, apparently distinctive, categories was criticised as overlapping. Allan (1992) in his validational studies of reading comprehension tests, gives a more detailed account of the degree of redundancy of these strategies.

In another research, previous to that of Anderson et al, Gordon (1987) designed a similar study. The aim of it was to empirically investigate the effect that testing method has on achievement on reading comprehension tests in EFL by the use of free-response and multiple-choice items. In addition to that, she examined the test-taking strategies that are employed in the two methods by means of introspection.

The results of her empirical data showed different indices for the two formats and after the analysis of the verbal protocols, she classified students' test-taking strategies as common to both testing modes and unique to each. One of the researcher's conclusions was that the "strategies which emerged from the qualitative data must now be validated empirically and on a broader population to determine the extent to which they apply to other learner populations" (ibid: 156).

Alderson (1990) in his attempt to get closer to the test-taking process of two EAP students while taking a reading test, consisting of free-response items, talks of a method effect. More particularly, he contrasted the reports of the warm-up phase, on a multiple-choice reading test used for the purpose of familiarising the students with the technique of the "think-aloud" report with that of the free-response reading items used for the study. What he noticed was that these two students used different patterns of

behaviour in their attempt to give answers to the questions. The researcher concluded that " the 'same' question in two different formats may very well involve test-takers using different processes or skills" (ibid: 470).

This was investigated further by Allan (1992) in his 5th study of a series of process-based studies examining the effect of item type and format upon readers' selection of test-taking strategies. In his study, 25 students at the City Polytechnic of Hong Kong were given a passage from a demonstration TOEFL examination paper to read. Fifteen of those were presented with six items in the original set of multiple-choice format and the remaining ten were given the same questions but rewritten in free-response mode and were asked to think aloud in English as they worked through the tasks, a procedure that they had practised a week earlier.

The aim of the study was to investigate (a) how far specific categories of questions call for particular strategies, and (b) the effect of test item format on the test-takers' processing of the task at hand. The protocols gathered from the students were analysed and matched against the TOEFL official publication (quoted in Allan, 1992) which indicated two things. The first was what each question was measuring and the second was an outline of one way of arriving at the correct answer, which the researcher designated as "the projected strategy".

The comparison of the process-data with these pre-constructed task analyses revealed some very interesting tendencies concerning the two formats under examination. These made the researcher conclude that the two "formats tend to engage qualitatively different test-taking processes in students" (ibid: 574) and that "particular questions tend to encourage different sets of strategies" (ibid: 427).

2. 5 The Use of Self-Report Instruments as a Means of Collecting Processing Data.

A relatively new concept in the attempt of the investigators of reading who want to collect on-line data from as a big a number of subjects as possible, is the use of self-report checklists of reading or test-taking strategies. Such instruments supply quick

data collection presented in a form almost ready for analysis. At the same time, they give the chance to investigators to avoid falling in the trap of the case study category and limited generalisability of their research.

The first attempt of using such a checklist is that of Groebel's (1981). The purpose of her investigation was to find out which reading techniques are recommended by EFL teachers and what is actually practised by the students. She administered it in the form of a small scale questionnaire to university students and their teachers. The questionnaire listed 15 reading techniques and asked the students to put them in the order in which they would normally use them. Teachers were instructed to put the techniques in the order which they believed students should use them when dealing with a reading passage. The results showed disagreement of opinions among the students and their teachers.

Unfortunately Groebel did not try to construct her questionnaire from baseline data gathered previously from similar respondents but she did so based on theory only which makes the results sound of suspect validity.

A second attempt to gather immediate report data in the form of a checklist was that of Nevo's (1989). Nevo, too, composed a self-report checklist of test-taking strategies "based on test strategies described in the literature and on personal intuitions as to possible strategies which respondents might select" (ibid: 204). This was administered along with a multiple-choice reading test. It was hoped that this research design would allow for immediate feedback after each item and thus get closer to the way students process multiple-choice reading items

The checklist - the researcher does not mention in which language it was written - comprised 15 strategies with an open 16th one, labelled "Other", contrary to Groebel's closed questionnaire. The main finding was that it was feasible to obtain feedback from respondents on their strategy used after each item on a test, if a checklist was provided.

Nevo's attempt was criticised by Allan (1992) for weak validity. The flaws the latter sees range from the way the administration was done to the instructions given to the students for strategy selection. More specifically, Allan claims that Nevo used "unconscious pressure" on the students by exemplifying test-taking behaviour and asking them not to skip items, "forcing" thus the students to do guesswork and to admit so later by selecting the corresponding strategy.

Furthermore, by requiring subjects to mark for each item they answered both the "primary" and "secondary" strategy, does not necessarily mean that they used two strategies. In addition to that, the instructions did not allow students to show the strategies they used and abandoned later in an attempt to reach their final choice. He goes on to say that it is not always possible to be able to discriminate with certainty which strategies are primary and which ones secondarily, a division that seems to have been done somehow arbitrarily. Results were mainly reached by a tabulation of the frequency of use of primary strategies and a discussion followed about which ones of these help the test-taker to reach the correct answer, while nothing is mentioned about the function or use of secondary strategies.

Thinking of a possible checklist effect that Nevo did not consider before launching her instrument, Allan (1992) attempted to validate the checklist. So what he actually did in his fourth study, was to look more closely at the invalidating effect of the same self-report inventory when he administered it to four different groups of students with similar characteristics under four conditions:

(a) complete list of strategies given,

(b) complete list of strategies reordered,

(c) list of strategies minus the most popular strategy and

(d) list of strategies minus the most popular strategy and the next most popular strategy.

Under a fifth condition, in which no checklist was given, another group of students were to report their mental processings while taking the same test to see if their verbal reports would resemble the strategies drawn in Nevo's instrument.

Frequency analysis of students selections showed reasonable evidence that this self-report checklist of reading and test-taking strategies influenced respondents in their choices of strategies. The researcher cautions for attention in the case of using a similar instrument, issuing a call for cross-validation against verbal report data.

Research has already indicated a method effect in reading comprehension tests and investigators have manifested the need to gather not only product but process data too, in order to establish construct and content validity of reading comprehension items. This study, as well, attempts to investigate the method effect of two testing techniques, namely free-response and multiple-choice formats, and to throw some more light on the cognitive activities test-takers employ when students interact with the experience of taking these reading comprehension tests.

CHAPTER 3

PURPOSES AND DESIGN OF THE PRESENT STUDY

3. 1. Purposes of the Study

This study mainly aims at exploring and examining the nature of test method effect by using both a product- and a process-based approach to EFL reading test validation. The methods selected are free-response format and multiple-choice format on the grounds of being so frequently employed by test constructors and teachers.

It was posited that if the two methods are measuring the hypothesised trait of reading comprehension then the test-takers would perform equally well on both tests of reading comprehension and would employ the same test-taking strategies. Test-taking strategies are defined here as "those activities in which a test-taker engages in order to provide an answer to the testing task" (Gordon, 1987: 82).

Two main hypotheses were generated:

Hypothesis A: There would be no significant difference between scores obtained on the free-response and multiple-choice reading comprehension tests.

Hypothesis B : The test-taking strategies employed in answering reading comprehension questions in free-response and multiple-choice formats would be equivalent in that they would be measuring the same trait.

To confirm or disconfirm these two hypotheses, two types of data were collected from students: information on their performance on the two reading tests and process data collected by means of a self-report Checklist of test-taking strategies.

The Checklist was also validated in two ways, in order to investigate any possible effect that the use of such an instrument might have on the responses of the students. According to the first one, the most popular strategy on the Checklist was deleted for half of the students on both administrations. The second way was by mapping strategies that two students used during their introspective interviews, onto strategies of the Checklist (the rational for the validation procedure of the Checklist is given later in this chapter).

The present study was designed also to take into account students' attitudes towards the two specific formats, look at more general reading strategies and find out whether students' perception of test difficulty matched their actual difficulty in performing the reading tasks.

3. 2 Research Design

3. 2. 1 The Materials

The two reading tests selected, were adapted from Allan (1992) who had used them for his fifth experimental study (briefly discussed in Chapter 2). They both consist of the same passage and are followed by the same 6 items in two different formats: free-response format for the first and multiple-choice format for the second. These reading items were designed to measure the same attributes in both tests (see Appendices A and B).

The passage along with the multiple-choice items, accompanied by four options each, were taken by Allan from *Understanding TOEFL*. This is an official publication for the TOEFL program produced as a sample paper for future candidates (Educational Testing Service, 1987a, quoted in Allan, 1992). The free-response items were written by the same researcher for the purposes of his study .

The passage is of expository nature, with a considerably high lexical condensity, which is a characteristic of scientific discourse. It discusses treatments for bee stings and introduces a new means of immunotherapy using bee venom. The selection of the particular reading test was felt appropriate since many of the students who took part in the present study, might want to pursue further education abroad and thus would have to take a similar reading test.

Understanding TOEFL Workbook (Educational Testing Service, 1987b, quoted in Allan, 1992) provides the specifications of what each question is designed to test. (An account of what each item measures, is given in **4. 2. 1.**)

3. 2. 2 Designing the Checklist of Test-taking Strategies.

Introspective interviews were conducted with a total of five native and non-native teachers of English, currently doing an M.A. in the Linguistics Department at the University of Lancaster, who voluntarily took one version of the test.

The teachers were asked to read the text, answer each question and to think aloud in English as they worked through the tasks. Their remarks were automatically recorded onto audio tape and notes were taken by the researcher on their general reading strategies which formed the basis of the Retrospective Questionnaires later (see Appendices A and B). They were already familiar with the think aloud technique, as all of them admitted having done something similar in the past, but not for a reading test.

The interviews were transcribed, reviewed and notes were made relating to processes involved in answering the questions. Following a data-derived analysis of the protocols, test-taking strategies were then elicited.

A Checklist of Test-taking Strategies was constructed in English, based on the verbal protocols of the teachers, reviewing the relevant literature and on personal intuitions about behaviour that might be elicited by similar items and question types. The Checklist was translated in Greek, the mother tongue of the target population, for reasons discussed later in the chapter and the translation was also checked by a Ph.D. Greek student.

3. 2. 3 Piloting the Instruments.

The two reading tests, along with the translated version of the Checklist, were piloted on 10 M.A. Greek students at the University of Lancaster, from a variety of disciplines.

The rational behind it was to check the clarity of the strategies in the Checklist but primarily to find out if there were any particular strategies that the students might favour and where these would appear on the Checklist. This is a point which was taken into account when the final versions of the Checklist were constructed.

In light of the feedback received, final modifications to test procedure, clarity of instructions and wording of strategies were made before administering these tests to the target population.

3. 2. 4 The Students.

Fifty-three female and four male Greek graduate students from the English Department of the University of Athens who were attending a two-week summer school at the Institute for English Language Education at the University of Lancaster, took part in the study. All subjects were speakers of Greek and had studied English as a foreign language for almost 14 years.

Their age ranged from 21 to 24 and almost all were in command of another foreign language in addition to English. They all had taken EFL examinations in the past, mainly the Cambridge Language Exams which are considered a prerequisite for academic and professional advancement in this country.

Before taking part in this study, they were told that its purpose was to find out about examinees' test-taking strategies while taking a reading comprehension test and two of them offered to take part in the introspective interviews (other than the 57 students). During the administration of the tests they were randomly divided in two groups of 29 and 28 students each, for reasons discussed in the following section.

3. 2. 5 The Instruments: Rational and Administration.

Four Instruments were devised for the study. These were administered with the following sequence: **Version A** (29 students) and **Version AI** (28 students), both in free-response format taken on the same day. A week later, students were given the same text but with the items presented in multiple-choice form, hereinafter referred to as **Version B** (28 students) and **Version BI** (29 students), both in multiple-choice format taken on the same day. The students took the free-response version first so as not to have the chance to look at any alternative solutions provided in the multiple-choice mode.

All Instruments were mainly comprised of four main parts, with the exception of **Version A** and **AI** which had attached to it one more section entitled *Personal Information* (see Appendix A) meant to collect biodata about the subjects.

Each of the four **Versions** comprised of four parts as follows:

(1) Preliminary Information: This part of the test included instructions on how to take the test divided in three steps which were kept almost the same in all versions (see Appendix A and B).

(2) Checklist of test-taking Strategies: The Checklist of specific strategies for immediate introspective use after each item, was to be read first and consulted later again after each item was answered. Each strategy was given a code number and a short description of one or two words, for easiness of reference. It was the only part in each Instrument that was written in Greek, the mother tongue of the students, so as not to interfere with the test-taking process imposing thus a second reading "task" that could have slowed down their process or be perceived as an extra "reading test".

The Checklist for each version was different in number. For **Version A**, the Checklist devised consisted of a total of 10 strategies (see Appendix C). The last one, **No 10. Other strategy**, refers to the use of a strategy other than the ones provided on the list, keeping it open to those students who felt that they had used a different strategy to put it down as No 10 and give, if possible, an explanation of this other strategy.

In consideration of a potentially invalidating effect upon readers' responses, the same Checklist, given in **Version AI** , with a total of 9 items this time, (see Appendix C), was given to the rest half of the students because of the following condition. If the most frequently cited strategy, **No 5. Locate**, (proved so during the piloting stage) was deleted from the list, students would tend to use the open-ended strategy, this time **No 9. Other strategy**, and would provide brief descriptions that would reflect the strategy deleted. If they did not do so and used some other strategy from the list which

seemed to cater for their mental processing, then the Checklist would appear to be exercising an effect on the test-taker. This would mean that either the list exhibits an amount of overlapping of strategies or is perceived as so by the students thus reducing the precision of the Checklist and invalidating it to an extent (a condition used in Allan, 1992).

As mentioned above, a week later, so as to minimise as much as possible the chance of remembering much from the first administration, students took the same text with 6 items presented in multiple-choice. The Checklist used in **Version B** (see Appendix C) consisted of 13 categories with the last one, **No 13. Other strategy**, left open. The only difference between this and the previous ones was that in some of the descriptions of the strategies the word *alternative* was added, taking thus into account the inherent characteristic of the multiple-choice format. In addition to that, three more strategies were added to this Checklist, that is **No 10. Match stem** , **No 11. Eliminate** and **No 12. Deduction** which were felt specific to this format.

On the same day the rest half of the students took **Version BI**, with the Checklist comprising 12 items (see Appendix C), since **No 5. Locate** was deleted again in an attempt to validate the Checklist for this administration. This immediately renders **No 12. Other strategy** as the open-ended strategy.

(3) The reading comprehension test: The materials used in this part have already been presented in **3. 2. 1.**.

What needs to be said here also, is that each reading item, whether in free-response or multiple-choice mode, was followed by an introspective part, first asking the subjects to indicate the number of strategy (-ies) they thought they had used on answering each item (see Appendix A and B). This was expected to be done by consulting the Checklist given.

If students felt they had used a strategy (-ies) other than the ones on the Checklist given, space was also provided for them to give a description of this other strategy (in English or in Greek).

Students were also asked to indicate the certainty of their choice of strategies on a three-point scale. Descriptors for each point were provided in Preliminary Information.

(4) Retrospective Questionnaire: This part, divided in three smaller ones, enquired about students' general reading strategies in taking the specific version, their estimation of the difficulty level of items and a more general reaction to the format currently assessed on (see Appendices A and B).

The following table gives a description of the administration of the tests over the period of the two weeks:

Table 3. 2. The administration of the tests along with the Checklists over the 2 weeks.

1st Week	*2nd Week*
Version A/AI (6 free-response items) (n = 57)	Version B/BI (6 multiple-choice items) (n = 57)
Version A Checklist : 10 strategies (n = 29) Version AI Checklist : 9 strategies (n = 28)	Version B Checklist : 13 strategies (n = 28) Version BI Checklist : 12 strategies (n = 29)

Students were told that they had 40 minutes maximum in their disposal during the first administration and 30 minutes for the second one. This time limit, which was decided during the piloting stage, was felt necessary so as to create real testing conditions and avoid as much as possible the contamination of data if students were allowed to take the test without the constraint of time limitations.

The students were also told that their individual test scores would not be revealed to anyone and that their test performance would not be used for any other purposes except by the researcher to determine how students take tests.

Students were advised to behave as they would have done on any other test-taking situation.

3. 2. 6 The Scoring Procedure.

All items were marked as either right or wrong. No half-marks were given. Intra-rater reliability was established by correcting all free-response items on two separate occasions allowing for an interval of almost ten days. The reliability index was 0.93 which is high enough to be relied on. For those items that there was disagreement, a joint decision was taken along with another marker, an M.A. student in the Department of Linguistics at Lancaster University, specialising in testing.

Multiple-choice items were also double-checked for any possible inaccuracies in scoring.

3. 2. 7 The Introspective Interviews.

3. 2. 7. 1 The Purpose.

A severe test on a possible influence of the Checklists would be also to present the same text and questions in free-response and multiple-choice formats to students who are similar to those who used the Checklists, and to elicit from them the strategies they would use to give answers to these items, without having access to any of the Checklists. If there is no similarity between the strategies which they report having used and those on the Checklists, it can be concluded that the instrument does affect test-taker responses and is therefore of suspect validity. For this cross-check two subjects were interviewed on their test-taking process.

3. 2. 7. 2 The Subjects, Materials and Administration.

Two subjects, belonging to the same group of students described above, volunteered to take part in this phase. They will be referred to as Subject A and Subject C in the following sections.

The material given to them was the same reading passage, as for the rest of the students, followed by the same six free-response and multiple-choice items. Both

subjects took the free-response version first and a week later, they took the multiple-choice version of the test on the same day as the rest 57 students.

3. 2. 7. 3 The Training Sessions.

To make the subjects aware of the elicitation procedure, and also to give them the opportunity to become acquainted with the researcher, two meetings took place earlier to the introspections.

During the first one, an informal conversation took place to establish rapport and to facilitate communication between the subjects and the present researcher. At the end of this session they were both asked to take a reading multiple-choice test.

Three days after the first meeting, each subject met with the researcher on different occasions and they were asked to look at the reading tests they had already taken and try to reflect on how and why they had responded they way they had. A reading test was preferred rather than a reading passage in order to give the two subjects the chance to practise reporting their test-taking strategies, so crucial for the present research.

3. 2. 7. 4 The Procedure followed during the Interviews.

During the first administration, the passage followed by the set of the six free-response items was given to the subjects. They were asked to express their thoughts in their mother tongue so as not to impose a further demand on their processing capacities and to collect as complete reports as possible.

They were also asked to report what they did after completing each item (immediate retrospection). On doing so, Subject A, after completing the first item, expressed the wish to verbalise her thoughts while doing the task (introspection), because she felt she did not "like to be interrupted".

Surprisingly enough, Subject C, did the same change during the second administration. Therefore, question probes, of an open-ended nature, were used only with Subject C during the first administration.

During these sessions the researcher would occasionally jot down notes describing the subjects' overt behaviour (i.e. reference back to the text).

3. 2. 7. 5 The Analysis of the Verbal Protocols.

During the interviews, the subjects were tape-recorded for later transcription. The sessions lasted on average thirty minutes and were later translated in English.

The protocols were analysed in the following two ways by the present researcher: Firstly, it was determined to what extent the two volunteers used words or phrases similar or identical to those in the Checklists. Secondly, these elements were mapped onto strategies in the Checklists and were labelled with the same code numbers as they appear in the Checklists (see Appendix E). The same procedure was also followed for the analysis of what the rest of the students wrote for **Other strategy** (see Appendix D).

The results would prove how many, if any at all, of the strategies already on the Checklists were also used by the two subjects during the interviews. If subjects reported having used same or similar strategies with the ones on the Checklists, this would further validate the selection of the strategies used.

CHAPTER 4

ANALYSIS OF THE RESULTS

This chapter consists of two sections. The first section will present the quantitative analysis of the students' performance on the two reading tests. The second will focus on the analysis of the process data obtained by means of the Checklists and will also give a report of what students said in the Retrospective Questionnaire.

4. 1 Quantitative Analysis of the Results

Item difficulty and item discrimination were calculated for all six items of both free-response and multiple-choice tests using the Microcat computer program ITEMAN. A complete breakdown of item statistics is given in the following table:

Facility Value and Discrimination Indices for both Free-response and Multiple-choice questions.

	Facility Value		Discrimination Index	
Questions	Free-response	Multiple-choice	Free-response	Multiple-choice
1	.59	.82	.64	.47
2	.89	.70	.53	.17
3	.98	.94	.22	.50
4	.52	.87	.59	.41
5	.31	.59	.49	.62
6	.78	.63	.49	.45
mean	.68	.76	.49	.44

It can be seen from the above table that items intending to measure the same trait were not necessarily of equal difficulty. This is more obvious with items 1, 4 and 5, which were more difficult in free-response than in multiple-choice form as opposed to items 2, 3 and 6 which yielded higher facility values when they appeared in free-

response format. The alternative answers presented in the multiple-choice format seem to have had facilitating effect over the first three of these items but appear to have confused students in the second case.

When items 1, 2, 4 and 6 appeared in free-response format, they discriminated the students better than in multiple-choice form. The easiest item in both formats was number 3 while the most difficult was number 5. Interestingly enough, both items yielded greater discrimination among students when they were in multiple-choice than in free-response format.

Another interesting variation between the two formats is seen in item 5. Despite its overall difficulty, when it appeared in free-response format 24.6 % of the students left it unanswered and wrote instead that the answer is nowhere to be found in the text. When the item appeared in multiple-choice form far less students abandoned the item - only 5.3 %. This supports the assumption that the alternative answers, which is characteristic of the multiple-choice format, have helped students to provide an answer which they could not do when the item was given in free-response form.

By comparing mean item facilities and discrimination indices for both versions, the multiple-choice test appears overall to be easier and to discriminate less well than its free-response equivalent. The means of the two tests, which were 4.6 for the multiple-choice test and 4.1 for its free-response counterpart, further support the fact that the multiple-choice version was easier.

A point worth noting here is that students' estimation of the difficulty level of the items conformed, in most cases, with their performance (see Appendix F). Based on the number of students who responded when asked about which of the items they found easy and which they found difficult on the Retrospective Questionnaires, it can be said that students were able to predict the level of difficulty for the multiple-choice items more accurately than for the free-response ones. For example, items 2, 3 and 4 in the free-response format were considered easy by the same number of students, while the facility values were actually different for each. This did not happen with the multiple-choice items whose difficulty index was correctly predicted by the students.

The reliability indices calculated for the two tests using KR20 turned out to be very low. More specifically, for the free-response format the reliability index was 0.43 and for the multiple-choice format it was 0.11

One major factor that appears to have yielded such low indices is the number of items in both tests (only six items in each) and the low discrimination indices for some of the items which students on the whole seem to have found very easy. This is even more obvious in the case of the multiple-choice items, which seem to have affected the correlation index for this test (see **Table 4. 1 .**).

This low reliability index could also be attributed to the sequence effect, that is, the students saw the text for the second time and it was short enough to be remembered. According to what the students said, this might be so, from seeing the correct answer.

Furthermore, six alternatives were non-functional as none of the students chose them. Probably this was due to the fact that it was difficult to find alternatives for such a short text. Additionally, some students did not provide answers to two of the questions in multiple-choice form (question 5 as discussed above and question 6, for which 1.8 % of the students did not provide an answer in both formats).

The internal correlation between the two tests was very low as well : 0.25 which might have been affected by the low reliability index and the small number of items on the two tests.

On the whole it can be said that the variation observed between the two tests purporting to test the same skills, provides evidence of a possible method effect which disconfirms Hypothesis A. At this point, this is inferable only from the analysis of psychometric data. This assumption needs to be further supported by looking at process data rather, i.e. the cognitive activities students were engaged in while taking the two tests, before any final conclusions can be drawn.

4. 2 Qualitative Analysis of the Results.

A note on how the data is presented

The following section is a presentation of the data collected by means of the Checklists and the Retrospective Questionnaires given along with the two reading tests. It is divided in three parts:

In the first part, the frequency of strategy use in the two formats is presented for all six questions separately, as can be seen in **Table 4. 2. 1.** on the following page (reference to it will be made throughout the presentation). A quotation from the official TOEFL publication (Educational Testing Service, 1987b, as quoted in Allan 1992: 339-381) is given at the beginning, explaining what the question is measuring. A comment follows the presentation of strategy frequency by first confirming or disconfirming Hypothesis B. Secondly, it determines whether the Checklist was a reflection of students' processings by looking at what its validational procedure had yielded.

Reference is also made to the following Appendices: Appendix D, which is a compilation of what students said when they felt they had used a strategy other than the ones on the Checklists; and Appendix E, which also gives an account of what Subjects A and C said about the strategies they used during the introspective interviews.

The second part is a more general discussion on the overall frequency of strategies as used by the students.

The third part reports on the information gathered by means of the Retrospective Questionnaire.

Table 4. 2. 1. Overall frequency of Strategy use of Free-response and Multiple-choice format for all six questions *

	QUESTION 1				QUESTION 2				QUESTION 3				QUESTION 4				QUESTION 5				QUESTION 6			
Strategies	A	AI	B	BI	A	AI	B	BI	A	AI	B	BI	A	AI	B	BI	A	AI	B	BI	A	AI	B	BI
Guess	4	3	5	2	1	1	2	3	0	3	1	3	3	1	2	3	2	0	1	0	2	1	3	0
BK knowl.	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	4	1	1	0	1	1
Chronolog.	1	5	1	1	7	14	2	6	2	10	2	12	1	4	2	5	2	6	1	6	1	2	1	2
The whole	5	7	5	4	1	5	2	3	1	5	2	3	5	5	2	3	7	4	6	6	1	4	3	6
Locate	14		6		22		11		25		18		13		16		11		10		14		12	
Match text	2	1	0	3	2	3	2	4	0	2	2	3	0	2	1	3	2	1	1	4	1	1	0	3
Memory	6	11	13	11	1	4	4	4	0	4	3	1	8	15	5	10	3	3	1	3	11	17	5	5
Clues	1	0	0	0	0	0	0	0	1	0	0	0	2	1	0	0	0	0	0	0	2	3	0	0
Return later	1	0	0	0	0	1	6	2	1	0	0	0	0	0	0	0	6	10	1	1	0	0	0	0
Other	2	4	1	1	0	2	0	1	0	5	0	6	2	2	1	2	1	2	1	2	0	2	2	2
Match stem			0	1			1	0			0	1			0	1			0	2			0	0
Eliminate			0	0			2	2			1	0			1	0			1	0			0	4
Deduction			4	9			2	6			0	2			0	5			3	7			8	13
Total	36	31	35	32	34	30	34	31	30	30	29	31	34	30	30	32	34	28	30	32	33	30	35	36

* The boxes in which there is no mention of strategy use, indicate that the strategy / strategies for this Checklist have been deleted.

4. 2. 1 Strategy Use per Question.

Question 1.

TOEFL WORKBOOK Comment: This question tests the main idea of the reading passage.

As can be seen in **Table 4. 2. 1.**, students used the same overall amount of strategies when this question appeared in the two formats - 67 occurrences in both cases. They selected most of the strategies with varying frequencies, with the exception of **BKknowl.** which was not used by any of the students across the Checklists. Students seem that they could not draw on any previous knowledge in order to answer this question. Probably this was due to unfamiliarity of the topic dealt in the passage.

Interestingly enough, when this item was presented in free-response form with Checklist **A**, students tended to use strategy **Locate** more often. When the same item was in multiple-choice along with Checklists **B** and **BI**, students reported having used strategy **Memory** more often. The selection of this strategy might have been due to the fact that students saw the text twice and therefore relied more on their memory, as they indicated.

When **Other strategy** was selected by the two students who used Checklist **A**, they expressed the need to return to the text again (see Appendix D). This might have been so because of "surface reading", as one of them admitted. Neither of them specified further what they did after that. When Checklist **AI** was administered, four students reported having used the open-ended strategy and the explanations provided by the first, third and fourth reflect the missing strategy **Locate**. Still the need to return to the passage again is obvious from what the second and third student said, in order to " make sure that the answer is correct", as the second said.

The protocols of Subjects A and C provide evidence of the most frequently used strategy **Locate**. In addition to that, they both confirmed what the previous students said for the open-ended strategy : the need to return to the passage again to look for appropriate words/phrases and to "confirm" their answers. This kind of behaviour can

be attributed to the format used, since students are required to provide the answer themselves when an item is in free-response format.

The **Other strategy** was used by only two students when the item appeared in multiple-choice form. More specifically, when Checklist **B** was used, the student who chose it explained that s/he had used the **Deduction** strategy and the student who used Checklist **BI** explained that s/he had actually used two : **Match text** and **Deduction** . On introspecting, Subjects A and C appear to have used **Deduction** as a way of giving an answer to this multiple-choice item.

Comment: On the whole, there seems to be evidence of possible method effect reflected in the choice of the most frequently used strategy between the two formats. For the free-response format it was **Locate** and for the multiple-choice format it was **Memory**. Another interesting difference is that almost a quarter of the students chose **Deduction**, the last of the three method specific strategies, when the question appeared in multiple-choice form that further supports the difference in the way students have processed this item .

The verbal protocols and the explanations of some of the students who used the open-ended strategy reflect strategies similar to the ones on the Checklists.

QUESTION 2

TOEFL Workbook Comment: This question tests a supporting idea.

Almost the same number of strategies was employed for this question, in the two formats. The strategies that were not chosen by any of the students for this question were **BK knowl.** and **Clues** . This comes contrary to the reassurance of Subject A, who said that she could see some link between this question and the answer to question 1, when she was introspecting on the item in free-response format (see Appendix E).

The most popular strategy in both intact Checklists **A** and **B** was **Locate** but with a slight difference: only 11 students chose it when the question appeared in multiple-

choice compared to 22 when it was presented in free-response format. Probably when the item appeared in multiple-choice form, students felt they needed to employ other strategies as well, in order to find the answer.

An interesting difference was observed here between the processing of the item in the two forms: 8 students altogether used strategy **Return later** as a way of providing an answer to this item when it was in multiple-choice mode but only 1 student felt s/he had done so when the item was presented in free-response form. Probably the provision of the alternative answers for the question in multiple-choice form must have confused these students who decided to reconsider the item later again.

None of the students chose the **Other strategy** when they were given Checklists **A** and **B**. But two students chose the open strategy when Checklist **AI** along with the item in free-response format were provided. The explanation of the first one reflects the missing **Locate** and the second seems to have given a rather vague description of how s/he processed the item but with a strong reference to the text again (see Appendix D). Subject A, from what she said, appears to have used a variety of processing strategies before she finalised her answer, i.e. **Clues**, **Locate**, **The whole** and as usual, went back to the text again to look for appropriate "terminology", as she admitted. From what Subject C said, it seems that she has used only strategy **Locate**.

Only one occurrence of **Other strategy** was reported by students who used checklist **BI**, that is the list with the one strategy missing. From what the student has written, it seems that **Guess** was his/her policy for this item in multiple-choice format but goes on to explain that this was due to the fact that one of the distractors was an unfamiliar word. Subject A reported using two strategies for this version, i.e. **Locate** and **Match text**, while Subject C used three, i.e. **Chronolog.**, **Match text** and **Deduction**.

Comment: The most popular strategy for this question in both formats was **Locate**. This was reasonable since the question tested a supporting idea in the text that required from the students to use this particular processing skill. But this was not done

with the same frequency and students tended to use other strategies as well so as to give an answer to this question, especially when it was presented in multiple-choice format. It could be said that this question has spread students over a wider selection of strategies when it appeared in multiple-choice form.

The verbal protocols and the explanations of some of the students who used the open-ended strategy reflect strategies that are similar to the ones on the Checklist.

QUESTION 3

TOEFL Workbook Comment: This question tests a supporting idea.

With the highest frequencies observed in all cases, **Locate** is the strategy used the most by student from the two intact lists **A** and **B**. Even when the strategy was missing from Checklists **AI** and **BI**, 10 out of 11 students altogether provided descriptions of strategies that reflect this strategy (see Appendix D).

Due to the demand perhaps of the item probably on students' processing (to locate specific information in the text), there were strategies that were almost wasted, as can be seen in **Table 4. 2. 1.** This means that **BK knowl.**, **Clues**, **Return later** and **Match stem** were used by only one student.

Subject A said that she had used **Locate** and **The whole** as ways of arriving at her answer when the item was in the free-response form and the first strategy only as a means of processing the item in the multiple-choice format. Subject C also reported having used **Locate** when the item appeared in both formats and also **Deduction** as an additional strategy for the item in the multiple-choice form (see Appendix E).

Comment: Due to length of the text most probable and the nature of the question, this item seems to have been processed in similar way in both formats by the students.

Moreover, the explanations provided by the students for **Other strategy** and what Subjects A and B said, reflect strategies from the Checklists.

QUESTION 4

TOEFL Workbook Comment: This question tests an inference that you should make from reading the last sentence.

Apart from **Locate** being the most popular selection again, this item managed to discriminate students in their choice of the remaining strategies to an extent. For example, 10 students altogether claim to have employed strategy **The whole** when the item was in free-response format. The use of this strategy contradicts the hope entertained by the test-designer that students will infer the answer just by looking at the last sentence.

In addition to the above selection, strategy **Memory** was chosen by 23 students altogether when the item was in free-response form compared to only 15 occurrences of use when the item was in multiple-choice form. **Clues** was used by three students for the item in free-response form whereas no students selected it while processing this question in multiple-choice format.

Two strategies were not used by the students at all: **BK knowl.** and **Return later**. It can be said that the answer to the question must have been clear to the students who felt they did not want to employ this "buying time" strategy.

Two students also reported having used a strategy other than the ones on Checklist **A**. The first one refers to the need to confirm his/her answer with the text but does not explain further and the second one decided to go back to the text because s/he was looking for the "same words / expressions".

When Checklist **AI** was given to the rest half of the students, two of them chose **Other strategy** but only one provided an explanation which is similar to the missing **Locate**. Subject A used a complex of strategies: **Clues**, **Locate** and **The whole**. Subject C reports strategy **Locate** and the need to "make sure" of her answer, as she said.

The student who chose the open strategy for Checklist **B**, appears to have used a strategy similar to the one on the list, namely **Deduction**. From the two who selected the **Other strategy** from Checklist **BI**, only one gave an explanation. This student

expressed the uncertainty of the choice of his/her answer which tried to justify by returning to the text again. When this item in multiple-choice was given to Subject A, she reported strategies **Deduction** twice and **The whole**. The first of these strategies was also used by Subject C.

Comment: Except from the non-discriminating use of strategy **Locate**, there is only slight evidence which confirms the method effect hypothesis in the way students have processed this item.

The protocols of Subjects A and C reflect strategies included in the Checklists as well. Only two of the students reported similar strategies when they used the open-ended strategy.

QUESTION 5

TOEFL Workbook Comment: This question tests supporting information that is presented in different parts of the reading passage.

Although the most frequently cited strategy for this item is **Locate** again, the frequencies were relatively low with 11 and 10 occurrences compared to the frequency with which this strategy was used for the rest of the questions. This frequency is justifiable since the question was designed to test information presented in different parts of the text.

An interesting tendency here is that 16 of the overall 57 reported having used strategy **Return later** when the question appeared in free-response format, compared to only 2 when the item was presented in multiple-choice form. Probably when the question appeared in multiple-choice form, it provided students with possible answers to choose from and did not want to abandon the item and return to it later. Instead they used their deductive reasoning more in order to chose one of the alternatives and thus provide an answer.

The only strategy that was not used at all, was **Clues**. Due, most probable, to the complexity of the question, students could not see any link between this question and the rest. Some of them even commented that they could not find it in the text.

One student who reported having used the open-ended strategy of Checklist **A**, said that s/he could not cope with the question at all, and decided to skip it. The two students who chose this strategy when they were given Checklist **AI** provided explanations that reflect the missing strategy. Subject A reports a cluster of strategies when the item was presented in free-response format: **Return later**, **Chronolog.** and **Clues** (twice), while Subject C reported a different strategy, that is strategy **The whole**.

No explanation of his/her strategy was given by the only student who said to have used **Other strategy** when Checklist **B** was administered with the item in multiple-choice form. Two of the students who used Checklist **BI** and reported the use of **Other strategy** , have either decided to skip it or in the case of the second one, felt had used three: **Locate**, **Match text** and **Deduction**. Subject A and C used a cluster of strategies as well. They both report of four strategies that are similar to strategies on the intact list **B** (see Appendix E).

Comment: Despite the similar overall frequency of strategies used by students when the item appeared in both formats, there was a considerable amount of differences in the frequency with which students selected strategies for this question.

The strategies that Subjects A and C as well as some of the students who chose strategies other than the ones on the lists, were similar in meaning to the ones on the Checklists.

QUESTION 6

TOEFL Workbook Comment: This question tests an inference.

Despite **Locate** being the most popular strategy again, it is worth looking at what strategy **Memory** yielded. This was used by an overall of 28 students who answered the item in free-response format in contrast to only 10 when the item appeared in multiple-choice form. Probably students chose a different way to process the item in multiple-choice form.

Another interesting thing here is the frequency of strategy **Clues** . This was selected 5 times when the item was presented in free-response format while no students did so when this appeared in multiple-choice. The use of the method specific strategy **Deduction** by 21 students when this question appeared in multiple-choice form (the highest frequency of all cases).

All this variation in the way strategies were used, allows room for the assumption that a possible method effect was being exercised while students were trying to choose a plausible answer.

None of the students used strategies **Return later** and **Match stem** . Probably because students by this time had scrutinised this short text enough to come up with an answer and did not need to resort to these two strategies.

No students used the open strategy when the intact Checklist **A** was given to them while two did so with Checklist **AI.** Their explanations reflect the missing **Locate**. Subjects A and C, said that they remember the text quite well to give an answer to this item and went back to the passage only to "have a look at the wording" as A said. This was reported by Subject C who used four strategies this time in addition to this.

Two students who were given Checklist **B** along with the items in multiple-choice selected the open-ended strategy. They reported the same strategies as the ones on the list. The other two who chose the open strategy from Checklist **BI**, described the missing strategy along with the need to return to the text for "verification", as the second characteristically said. As for Subjects A and C, their protocols reflected strategies from the list, with a frequency of 4:2.

Comment: Along with the difference in overall frequencies (63 for the free-response format and 71 for the multiple-choice), this item has discriminated the students on their choice of strategies considerably.

From what the students said about the open strategy and from the analysis of the verbal protocols, it can be concluded that their strategies matched those on the Checklists thus validating its construction.

4. 2. 2 Overall Frequency of Strategies.

This section will refer to the overall frequency of strategies between the two formats and briefly comment on it.

Overall frequency of strategy use between the two formats.

	Free-response Checklists: **A + AI**	Rank Order	**Multiple-choice** Checklists: **B + BI**	Rank Order
Guess	21	**6**	25	**7**
BK knowl.	4	**10**	7	**11**
Chronolog.	55	**3**	41	**5**
The whole	50	**4**	45	**4**
Locate	99	**1**	73	**1**
Match text	17	**8**	26	**6**
Memory	83	**2**	65	**2**
Clues	10	**9**	0	**13**
Return later	19	**7**	10	**10**
Other strategy	22	**5**	19	**8**
Match stem			6	**12**
Eliminate			11	**9**
Deduction			59	**3**

The frequency with which strategies were used by students for the two formats seems to vary. Three strategies, as highlighted in the table above, have been ranked similarly by the students for both formats. The most frequently used strategy, in both formats, is **Locate**, which matches the findings of Nevo (1989) and Farr (1990), although they were both working exclusively on multiple-choice items. The later characteristically concluded that:

> " ..... the common element that directed the subjects was the focus on getting to the questions as quickly as possible and then using the questions to direct a search of the passage to locate the best possible information to answer the questions ". (ibid: 221)

Students here have used a similar way to process the items in both formats which is reasonable, since they are asked to take part in a test and therefore to answer questions.

On the other hand, the exact number of occurrences is substantially reduced for the multiple-choice format. This could be attributed to the fact that students in their effort to answer the multiple-choice items, employed other strategies form the Checklist as well, i.e. the three method specific strategies.

In addition to that, the third most frequently cited strategy for the free-response version of the test is **Chronolog.**, while for the multiple-choice version is strategy **Deduction**. This seems to disconfirm Hypothesis B, according to which the test-takers would use test-taking strategies that would be equivalents for both formats since the trait under examination is common.

Furthermore, 10 students selected strategy **Clues** when the items appeared in free-response mode while no students have resorted to the use of this strategy when the items were offered along with alternatives to choose from. A similar discrepancy can be seen for strategy **Return later** with 19 occurrences when the items were given in free-response format in contrast to 10 when these were written in multiple-choice form. It can be concluded here that when the questions were in free-response format, more students felt that they needed to look for clues from their answers to other questions or abandon their attempt to give an answer for a while and try again later. Free-response items seem, therefore, to exercise a different demand on the students. This comes contrary to the structuring effect of the multiple-choice format, that does not ask for the employment of these strategies so often.

Of interest is another frequency. Strategy **Match text** was chosen 26 times when the items were given in multiple-choice form compared to 17 instances in free-response form. This shows that when the students are engaged in a multiple-choice task, they required from the students to do a lot more "matching" between the items and the text while the students seem to have been engaged less in this activity when the items were given in free-response format.

The strategy that was used the less in both formats was **BK knowl.** . This means perhaps that students lacked any previous schema that they could resort to concerning the topic of the passage. Had the text been of a different topic, i.e. more neutral, it is possible that students could have selected this strategy, as well.

From the three method specific strategies that were added to Checklists **B** and **BI** the one that was used the most was **Deduction** with a frequency of 59 instances. This means that students were involved in this mental processing more often than the other two. On the other hand, strategy **Match stem** was used only six times altogether. This indicates that students were not involved so much in matching the stem sentence with the alternative answers but rather with the text, as mentioned earlier.

Contrary to the use of similar checklists where no students would chose the **Other strategy**, there was a considerable amount of students who not only chose the open-ended strategy but also described it. When these descriptions were given for Checklists **AI** and **BI**, they reflected the deleted strategy in most cases.

It is worth noting here that when students chose **Other strategy** they also expressed the need to return to the text in order to find "clues" for their answers or to "verify" the answers they had already given. This was more frequents when the items were presented in free-response form than when these were in multiple-choice format. It seems then that the students had to "plough" through the text more in order to find the "clues" they needed for the answers when the items were in free-response format.

To sum it up, the overall use of strategies for the free-response format was 380 while for the multiple-choice form this was 387. Despite the addition of three more strategies to Checklists **B** and **BI**, the demands of both formats on the students were of almost equal gravity and probably students felt they were using more than one strategy to find answers to these questions which yielded a high overall frequency for both formats.

This last point, though, needs to be further exemplified. As it can be seen from **Table 4. 2. 3.**, there was a considerable amount of instances where students reported they had used more than one strategy to give answers to the questions.

Frequency of more than one strategy per question

	Quest.1	Quest. 2	Quest. 3	Quest. 4	Quest. 5	Quest. 6	Total
A+AI	8	6	3	7	6	5	**35**
B+BI	8	7	3	4	4	12	**38**

Students have selected two, three and on some occasions four strategies in order to give answers to the questions. The selection of more than one strategies varies according to questions. Unfortunately the combinations of strategies students used for both formats did not reveal any particular patterns that could be further analysed and thus get a deeper insight into the way they would use these test-taking strategies. The selection seems to depend on individual students. Had a bigger population been used, it might have been possible to get a more concrete combination of strategies and be able to see further the requirements that the testing method or item content have on students' processings.

The above table shows that for questions 4 and 5, students reported that they had used more strategies when these items were in free-response format. When items 2 and 6 appeared in multiple-choice form, students felt they needed more strategies to handle the demand made on them by this format. This is also reflected in the facility values that these items yielded. It seems, therefore, that students needed more strategies to cope with items that proved to be more difficult when they were in one format rather than the other.

4. 2. 3 A General Comment on the Validity of the Checklist.

To determine to what extent the Checklist was a valid measure of the students' processings, as exemplified in the previous chapter as well, a double check was attempted.

Firstly, the verbal protocols of two of these students, were analysed (see also Appendix E). The explanations they gave about how the processed the items on the

two tests, reflected strategies already found in the Checklist. The second severe test on the instrument was when the most frequently cited strategy on the Checklist (determined so during the piloting stage) was withheld from half of the students. It was expected that since it was the most popular strategy, these students would use the open-ended strategy and would provide explanations that would be similar in meaning to the one missing.

It was found that not all of these students used the open-ended strategy but only a small number. When their explanations were analysed it was found that 22 of the overall 41 who chose this strategy, were able to give explanations that reflected the missing strategy. Although the number is not high enough, it is at least suggestive that these students were aware that the Checklist was not reflecting the strategy they used and tried to do so on their own. A considerable amount of metacognitive awareness is certainly needed in order to look back at the way(s) one has followed in order to give an answer to a question and to try and describe it. This requires training and willingness on both the behalf of the teacher and the student and depends entirely on individuals whether they are ready to become engaged in such a demanding task.

Surprisingly enough, students used strategy **Guess** for both formats with almost the same frequency although it is considered a strategy that testees would mainly resort to when taking multiple-choice tests. Probably they must have misunderstood its explanation. Had some explanations of the strategies given at the beginning this might have been avoided.

But the fact that a considerable amount of these students were able to identify strategies that they felt they were missing from the Checklist, comes contrary to the findings of other research done before on the Checklist effect (cf. Allan, 1992). This could indicate two things: either these students had the metacognitive awareness and appropriate language to specify their strategies or that a possible backwash effect was present here, since they come from an environment that test-taking behaviour is nurtured. Further research is definitely needed before anything can be said for sure.

4. 3 General Reading Strategies and Students' Perceptions of the Two Formats.

When students were asked in the first part of the Retrospective Questionnaires (see Appendices A and B) how they approached the text itself, based on the answers of those who responded to this part, it was found, that there were no significant differences in the way they would read the text or the questions in the two formats. This means that students tended to read the whole of the text first and answered the questions afterwards in chronological order.

This comes contrary to previous research. Li (1992) reports of a different reading procedure employed by his students who took a test with a heavy load of reading and more items in free-response format. The researcher noticed that more students began to answer the questions without first reading the whole text. Statistical analysis of the results showed that these students achieved higher scores compared to those that read the whole of the text first.

When students were asked to indicate which item they considered unfair they would mostly cite questions: 5 and 6 when these were in free-response format while they would add question 2 to that list when the items appeared in multiple-choice format.

Contradictory opinions were given when students were asked how they feel about reading language tests. Language was not a problem for the students who did not feel they needed to provide any of their answers in their first language although they were given the chance to do so.

At the end of the second administration, the students were asked to indicate which of the two formats they had found the most objectionable. 54.3 % chose the multiple-choice format and 43.8 % the free-response form. The rest of the students could not decide between the two. Furthermore, when they were asked to indicate on which of the two tests they thought they had performed most successfully, it was found that 56.1 % performed most successfully on the test they found most preferable and 70.1 % accurately predicted on which test they had achieved the highest score.

It is possible that the students' preference and results on the multiple-choice test were also influenced by the sequence effect of having seen the passage twice. This is brought out by the fact that 87.7 % of them thought that seeing the text twice has helped them to understand it better and 70.1 % believed that this also had affected their final score.

Almost all students admitted that the provision of options to choose from gave them ideas that they had not thought of previously but only 28 % of those who said that this had a facilitating effect performed better on these items.

30 % of the students expressed some uncertainty of their choice of strategy for question 4 and this was so when the item appeared in free-response format. This item also turned out to be more difficult than was expected (F.V. 0.52 although 40 out of 50 students classified it as an easy item).

On the whole, the students found the experience of reporting on their test-taking strategies as one of the first opportunities they had to find out more about how "the mind works when taking a test" and felt they had become more aware of "strategies and methods that previously remained unconscious" and "just a personal matter".

CHAPTER 5

GENERAL DISCUSSION OF THE RESULTS

The main objective of this study was to see whether the method of testing reading comprehension influences students' performance on the test and if it makes qualitatively different demands on their mental processings.

The results of the statistical analysis of students' performance on the free-response and multiple-choice tests indicated that there was slight evidence to support Hypothesis A.

The considerable difference of item performance between the two tests designed to test the same skills indicated that students did not perform equally well on the two tests. The items behaved differently yielding facility values and discrimination indices that were not the same over the two formats. For example, items 1, 4 and 5 proved to be easier in multiple-choice format than in free-response form as opposed to items 2, 3 and 6 which yielded higher facility values in free-response form. The effect of the alternative answers presented in multiple-choice appears to have been facilitating in the first case while they seem to have had a confusing effect in the second.

In addition to the above, the discrimination indices obtained for the items over both formats showed that two reading tests with identical content but different format can discriminate students differently. It can be said then that although these two formats indented to measure the same trait, they have actually yielded measures of different abilities.

This was further investigated by analysing the process data received by the students who retrospected on their test-taking strategies immediately after answering every item on the two tests by means of a self-report Checklist. The analysis of the frequency with which students chose strategies for the same questions in both formats was not similar. Nor was the overall frequency of strategies between the two methods.

This disonfirms Hypothesis B, according to which the test-taking strategies employed in answering reading comprehension questions in free-response and

multiple-choice formats would be equivalent in that they would be measuring the same trait.

More specifically, the overall frequency of strategy use between the two testing methods has revealed some very interesting patterns of behaviour. The strategy that was used most frequently by the students was returning to a specific part in the text to find clues in order to give answers to the questions. But this was done with greater frequency for the free-response items which actually require from the students to use their productive skills and provide the answer themselves, thus leading test-takers to plough more through the text for the discovery of appropriate words or phrases.

On the other hand, there was a considerable amount of students who said that they did not need to return to the text again after they had read it but rather relied on their memory of it so as to give answers to the questions in both methods. This contradicts the above and is quite revealing in the sense that not all students can be expected to follow the same mental paths in order to give answers to the questions. It would have been interesting to see if students would have relied so much on their memory had a longer text been used.

Multiple-choice items appear to be making an additional different demand on the students. That is, they require from them to use their deductive reasoning in order to eliminate the distractors and choose the appropriate answer. In other words, test-takers have to use their evaluation skills in order to identify which of the alternatives presented for choice is the correct one. In addition to that, far more matching of information between elements in the text and the question/alternatives was involved when students were working on the multiple-choice items.

Contrary to above, when the reading items appeared in free-response format, the test-takers resorted to sources that were not specifically related to the text in order to find an answer. Students felt that by returning to their answers in previous questions they might be able to find possible clues that would help them to provide an answer. Abandoning the item for a while and returning to it later, was another way that

students used in order to cope with items. This "buying time" strategy was used more often with items in free-response format.

It can be concluded then that students have interacted differently with the experience of taking free-response and multiple-choice reading items confirming the assumption that these two formats engage qualitatively different test-taking processes in examinees.

The Checklist proved to be quite reflective of the students' processing activities when it was withheld from the two students who thus reported their strategies unaided. Subjects A and B reported strategies that matched strategies that were included in the Checklist, validating thus its composition to an extent.

The second test of validity on the Checklist was when the most frequently cited strategy was deleted from it for half of the students on both administrations. Only a small number of students overall chose the open-ended strategy but the students' explanations were, in a lot of cases, similar in meaning to the one that was missing.

Students also tended to chose strategies from both the top and the bottom half. This is more evident when students were working on the multiple-choice items. They tended to use strategies from the bottom of the list as well. Therefore, there was not enough evidence to support the assumptions that there was a position effect which was exercised on the students' responses.

On the whole, it can be said that the Checklist devised, proved to be quite an efficient methodological tool that reflected the students' mental processing. Had a longer list with more items been devised, it could have been possible to cater for more subtle cognitive aspects in the test-taking experience. Further research, of course, needs to be done to determine which of these test-taking strategies can contribute to correct answers as opposed to incorrect ones so as to improve the quality of tests based on these formats and at the same time guide and train students in coping effectively with these test methods.

Analysis of the process data revealed an interesting behavioural tendency. Test-takers in order to arrive at their answer, seem to employ not only one mental

processing but more than that, the number of which is difficult to specify and probably depends on individuals, demands of the text used and item content. As one of the students said characteristically "different strategies or combinations of strategies can be applied in each question in order to obtain a correct answer". A complex of strategies or rather a repertoire was reported in several cases by students when choosing strategies from the list but also during the introspective interviews with the two volunteers. It definitely requires more researching on the processes involved in the test-taking experience in order to be able to see more clearly what particular strategies a test-taker employs in order to give answers to reading items for all methods.

In line with the above, the testee has to be recognised as a partner in the testing process whose judgement of task validity is worthy of consideration. This of course requires testee training but in the long run it could prove of benefit to both the test-designer and the test-taker.

The above findings have important implications for language testing and consequently for language teaching. Since tests are administered in order to obtain information about students, it is necessary to know precisely what is being tested. Thus, the method effect as well as the trait under examination needs detailed specification. In other words, if one chooses a method such as free-response, it is important to understand to what extent one is testing other traits as a result of the method employed. Similarly, the type of reading text as well as the item type should also be considered when choosing a testing method and these must be seen in reference to the particular purpose of the test. These and other factors that make up method effect therefore need to be specified for each and every method that is used in testing.

Though the main aim is to minimise method variance through the development of improved testing techniques, it is unlikely that any method of testing reading comprehension will completely eliminate its effect. Therefore method effect must be recognised and taken into account whenever assessment of students' abilities comes into play. What we can do is at least try and find more about its nature and the demands it makes on examinees.

One way of learning more about its nature is to obtain information on language learner's processes by the use of qualitative data. The information gathered in this study through qualitative modes of investigation has provided valuable insights in the processes test-takers engage in while attempting to provide answers to comprehension questions on reading tests. However, researchers, need to be aware of the possible shortcomings and limitations of such methods and also be careful when trying to overgeneralise their findings. By combining theoretical hypotheses with empirical and qualitative modes of investigation, we could possibly obtain a broader perspective on unobservable processes which hopefully will lead to a more in-depth understanding of what really is involved in the test-taking and reading process.

CHAPTER 6

CONCLUSION

This investigation has approached the use of various data sources in the process of examining the construct validation of two reading comprehension tests with identical content but different formats. The analysis of the data revealed that different formats intending to assess the same trait may yield measures of different abilities. It is therefore necessary to define as closely as possible the attributes of each method and take them into account when devising reading comprehension tests so as to control their effect on examinees' performance. This is directly linked to the needs of test desingers and teachers who want to be clear as to what their tests actually measure and what "intervening" factors they introduce by the employment of one method over the other.

Obtaining qualitative data on examinees' test-taking strategies can provide insights into the degree to which a specific testing method has affected performance. The means to collect such data can be varied and their reliability needs to be examined in order to determine the extent to which these have proved appropriate reflections of the test-taking experience. In addition to that, it was shown that a combination of more than one sources of data is needed in an attempt to gain greater insights into the reading comprehension process as well as the test-taking process.

The list of strategies in this study does not make claims of a pedagogical inventory but is rather exploratory in nature and open to interpretation. Further research needs to be done in order to be able to specify the delicate and at the same time complex cognitive activities involved in test-taking so as to establish a closer fit between what is really tested and what was purported to be tested.

The study has also shown that the role of the test-taker and his perceptions are of significant importance and need to be taken into account when attempts to construct validity are made.

Of course this study had its own limitations in the sense that the criteria for determining the strategies that appeared on the Checklist devised were not externally

validated. This means that before strategies on the Checklist were finalised, more than one opinions should have been sought especially when verbal protocols were analysed in search of the above mentioned strategies. Had this been done, the Checklist would have probably reflected students' processes to a greater extent and have exhibited less redundancy. Similarly, one more rater should have been involved in the categorisation of students' responses for both the open-ended strategies and the introspective interviews with the two subjects.

The test used turned out to be generally easy for this particular population which might have also affected their selection of strategies. It would be interesting to see what their choices would be if a text with a heavy demand on reading was provided. The fact also that they were exposed to the same short passage twice might have also affected their selection of strategies for the second test and inflated their overall scores. The types of information tested did not cover a wide spectrum and therefore the results can not be generalisable over more skills. The topic of the passage might have yielded results that were affected by its nature as well.

All this should be taken into account when the results are encountered and definitely need to be further investigated to determine the extent to which they are applicable to different age groups with varying language abilities.

BIBLIOGRAPHY

Afflerbach, P. and **Johnston, P.** 1984: "On the use of verbal reports in reading research". *Journal of Reading Behavior*, Vol. 16, 307-322.

Alderson, J. C. 1983: "The Cloze Procedure and Proficiency in English as a Foreign Language". In Oller, J. W. Jr (ed) "*Issues in Language Testing Research*". Rowley, Massachusetts: Newbury House, 205-217.

Alderson, J. C. and **A. H. Urquhart** 1988: "This test is unfair : I am not an economist", in Carrell, P.L., J. Devine and D.E. Eskey (eds.), "*Interactive Approaches to Second Language Reading*", Cambridge: Cambridge University Press, 168-182.

Alderson, J. C. 1990: "Testing Reading Comprehension Skills (Part Two)". *Reading in a Foreign Language*, Vol. 7/1, 465-503.

Allan, A. I. C. G. 1992: "*EFL Reading Test Validation: Investigating Aspects of Process Approaches*". PhD Thesis, Department of Linguistics and Modern English Language, University of Lancaster.

Anderson, N. J., Bachman, L., Perkins, K., and **Cohen, A. D.,** 1991: "An exploratory study into the construct validity of a reading comprehension test: triangulation of data sources". *Language Testing*, Vol. 8/1, 41-66.

Bachman, L. F. and **Palmer, A. S.** 1981: "The Construct Validation of the FSI Oral Interview". *Language Learning*, Vol. 31/1, 67-85.

Bachman, L. F. 1985: "Performance on cloze tests with fixed ratio and random deletions". *TESOL QUARTERLY*, Vol. 19/3, 535-556.

Bereiter, C. and **Bird, M.** 1985: "Use of thinking aloud in identification and teaching of reading comprehension strategies". *Cognition and Instruction*, Vol. 2/2, 131-156.

Block, E. L. 1986: "The Comprehension Strategies of Second Language Readers". *TESOL QUARTERLY*, Vol. 20/3, 463-494.

Bracht, G. H. and **Hopkins, K. D.** 1970: "The commonality of essay and objective tests of academic achievement". *Educational and Psychological Measurement*, Vol. 30, 359-364.

Cavalcanti, M. 1982: "Using the unorthodox, unreasonable verbal report protocol technique: quality data in foreign language reading research" in Dingwall, S., Mann, S. and Katamba, F. (eds), *Methods and Problems in Doing Applied Linguistic Research*. Lancaster: Department of Linguistics and Modern English Language, University of Lancaster, 72-85.

Cohen, A. D. and **C. Hosenfeld** 1981: "Some Uses of Mentalistic Data in Second Language Research". *Language Learning*, Vol. 31/1, 285-313.

Cohen, A. D. 1984: "On taking language tests: what the students report". *Language Testing*, Vol. 1/1, 70-81.

Cohen, A. D. 1986: "Mentalistic Measures in Reading Strategy Research: Some Reading Findings". *English for Specific Purposes*, Vol. 5/2, 131-145.

Dollerup, C., Glahn, E., and **Hansen, C. R.** 1982: "Reading Strategies and Test-Solving Techniques in an EFL-Reading Comprehension Test: A Preliminary Report". *Journal of Applied Language Study*, Vol. 1/1, 93-99.

Ericsson, K. A. and **Simon, H. A.** 1980: "Verbal Reports as Data". *Psychological Review*, Vol. 87/3, 215-251.

Farr, R., Pritchard, R. and **Smitten, B.** 1990: "A description of what happens when an examinee takes a multiple-choice reading comprehension test". *Journal of Educational Measurement*, Vol. 27/3, 209-226.

Frederiksen, N., Ward, W. C., Case, S. M., Carlson, S. B., and **Samph, T.** 1981: *"Development of methods for selection and evaluation in undergraduate medical education"*. Princeton, NJ: Educational Testing Service.

French, J. W. 1965: "The relationship of problem-solving styles to the factor composition of tests". *Educational and Psychological Measurement*, Vol.25/1, 9-28.

Gordon, C. M. 1987: "*The Effect of Testing Method on Achievement in Reading Comprehension Tests in English as a Foreign Language*". Unpublished M.A. thesis, Tel-Aviv University, Tel-Aviv, Israel.

Groebel, L. 1981: "The Students' Approach as Compared to Their Teachers' Recommended Approach". *English Language Teaching Journal*, Vol. 35/3, 282-287.

Grotjahn, R. 1987: "On the Methodological Basis of Introspective Methods". in Faerch, C. and Kasper, G. (eds), *"Introspection in Second Language Research"*, Clevedon: Multilingual Matters.

Henning, G. 1983: "Oral Proficiency Testing: Comparing Validities of Interview, Imitation and Completion Methods". *Language Learning*, Vol. 33/3, 315-332.

Klein-Braley, C. 1983: "A cloze is a cloze is a question". In Oller, J. W. Jr. (ed) *"Issues in Language Testing Research"*, Rowley, Massachusetts: Newbury House, 218-228.

Lewkowicz, J. A. 1983: *"Method Effect on Testing Reading Comprehension: A Comparison of Three Methods"*. M.A. thesis, Lancaster: Department of Linguistics and Modern English Language, University of Lancaster.

Li, W. 1992: "What is a Test Testing? An Investigation of the Agreement Between Students' Test Taking Processes and Test Constructors' Presumptions". M.A. thesis

Lancaster: Department of Linguistics and Modern English Language, University of Lancaster.

Mann, S. 1982: "Verbal reports as data: a focus on retrospection" in Dingwall, S., Mann, S. and Katamba, F. (eds), *Methods and Problems in Doing Applied Linguistic Research.* Lancaster: Department of Linguistics and Modern English Language, University of Lancaster, 87-104.

Nevo, N. 1989: "Test-taking strategies on a multiple-choice test of reading comprehension". *Language Testing,* Vol. 6/2, 199-215.

Oller, J. W. Jr. 1979: "*Language Tests at Schools*". London: Longman Group Ltd.

Olshavski, J. E. 1977: "Reading as problem solving: an investigation of strategies". *Reading Research Quarterly*, Vol. 12/4, 654-674.

Patterson, D. G. 1926: "Do new and old type examinations measure different mental functions?". *School and Society,* Vol. 24, 246-248.

Pritchard, R. 1990: "The effects of cultural schemata on reading processing strategies". *Reading Research Quarterly*, Vol.25, 273-295.

Sarig, G., 1987: "High-Level Reading in the First and in the Foreign Language: Some Comparative Process Data" in Devine, J., Carrell, P. L., and Eskey, D. E. (eds), "*Research in Reading in English as a Second Language*", Washington, D.C.: Teachers of English to Speakers of Other Languages.

Shohamy, E. 1983: "The stability of oral proficiency assessment on the oral interview testing procedures". *Language Learning*, Vol. 33/4, 527-540.

Shohamy, E. 1984: "Does the testing method make a difference? The case of reading comprehension". *Language Testing,* Vol. 1/2, 147-170.

Shohamy, E., Reves, T. and **Bejarano, Y.** 1986: "Introducing a new comprehensive test of oral proficiency". *ELT*, Vol. 40/3, 212-221.

Traub, R. E. and **Fisher, C. W.** 1977: "On the equivalence of constructed response and multiple choice tests". *Applied Psychological Measurement*, Vol. 1/3, 355-369

Vernon, P. E. 1962: "The determinants of reading comprehension". *Educational and Psychological Measurement*, Vol. 22/2, 269-286.

Ward, W. C., Frederiksen, N., and **Carlson, S.** 1980: " Construct validity of free-response and machine-scorable forms of a test". *Journal of Educational Measurement*, Vol. 17/1, 11-29.

Ward, W. C. 1982: "A comparison of free-response and multiple-choice forms of verbal aptitude tests". *Journal of Educational Measurement*, Vol. 6/1, 1-11.

TEST METHOD EFFECT

VERSION A / AI

PERSONAL INFORMATION

*Please fill in the following with information about you in **CAPITAL LETTERS**:*

Name (in full) : .....

Sex : .....

Age : .....

Foreign Language(s) (other than English) .....

Background of English Education (fill in the appropriate blanks) :

1. Private Language School: ..... years.
   (Frontistirio)

2. Public School: *Primary: ..... years.

   *Secondary: ..... years.

3. Private School: *Primary: ..... years.

   *Secondary: ..... years.

4. Private tuition: ..... years.

American or English Language Tests I have taken (e.g. FCE, CAE, CPE, TOEFL, IELTS .....):

.....

.....

Preliminary information

This is a test of reading comprehension. Before taking it, please **read all the instructions carefully.**

While taking the test, you will do the following **three** things:

Step 1: Answer the Questions in the spaces provided.

Step 2: When finishing answering **each** question, decide, based on the list of *Strategies* (given on the following page), which *Strategy* you have used in order to give an answer to this particular question. Beside each question write down the number of the *Strategy* which you used to answer it. If you used more than one *Strategy*, add that too.
If the *Strategy* that you used for a particular question is not on the list, then next to number " 10./9. Other Strategy " specify what other *Strategy* you used by giving a brief explanation (in English or in Greek).

Step 3: After finishing Step 1 and Step 2 , please rate how certain you are of the *Strategies* you have used to answer the questions. In doing this, you only need to circle one of the three numbers provided at the end of every question. What follows is the description of what every number stands for.

1. = I am certain that I have used the *Strategy (-ies)* I specified in Step 2.

2. = I am somewhat certain that I have used the *Strategy (-ies)* I specified in Step 2.

3. = I am uncertain that I have used the *Strategy (-ies)* I specified in Step 2.

Please keep in mind the following : If you make any changes in Step 1, that is, if you change your answer for the comprehension question, remember to change Steps 2 and 3, as these are affected by your previous change as well.

Read the following passage and answer the questions providing at the same time the information requested in Steps 2 and 3. You can <u>underline</u> or <u>keep notes</u> if you want. Please write your answers in CAPITAL LETTERS

A team of researchers has found that immunizing patients with bee venom instead of with the bees' crushed bodies can better prevent serious and sometimes fatal sting reactions in the more than one million Americans who are hypersensitive to bee stings. The crushed-body treatment has been standard for fifty years, but a report released recently said that it was ineffective. The serum made from the crushed bodies of bees produced more adverse reactions than the injections of the venom did.

The research compared results of the crushed-body treatment with results of immunotherapy that used insect venom and also with results of a placebo. After six to ten weeks of immunization, allergic reactions to stings occurred in seven of twelve patients treated with the placebo, seven of twelve treated with crushed-body extract, and one of eighteen treated with the venom.

1. What is the main topic of the passage?

.....

.....

.....

(Refer to <u>Step 3</u> and <u>Step 2</u> of the Preliminary Information Section if you are not sure of how to complete the following part).

The Strategy or Strategies I have used. Number(s): .....
If Number "10. / 9. Other Strategy", please specify (in English or in Greek) : .....

.....

.....

*Please <u>circle</u> **one** of the three numbers below to indicate the degree of certainty of your choice of Strategy Strategies :*

1. *2.* *3.*

BEST COPY AVAILABLE

2. What opinion do researchers have of the traditional treatment for bee stings?

.....
.....
.....

The Strategy or Strategies I have used, Number(s) : .....
If Number "10. / 9. Other Strategy", please specify (in English or in Greek) : .....
.....
.....

*Please circle **one** of the three numbers below to indicate the degree of certainty of your choice of Strategy Strategies : 1. 2. 3.*

3. How many patients took part in the experiment?

.....
.....
.....

The Strategy or Strategies I have used, Number(s) : .....
If Number "10. / 9. Other Strategy", please specify (in English or in Greek) : .....
.....
.....

*Please circle **one** of the three numbers below to indicate the degree of certainty of your choice of Strategy Strategies : 1. 2. 3.*

4. What was the most successful treatment described in the passage prepared from?

.....
.....
.....

The Strategy or Strategies I have used, Number(s) : .....
If Number "10. 9. Other Strategy", please specify (in English or in Greek) : .....
.....
.....

*Please circle **one** of the three numbers below to indicate the degree of certainty of your choice of Strategy Strategies : 1. 2. 3.*

BEST COPY AVAILABLE

5. In order to be successful, how must the treatment referred to in the passage be administered?

.....

.....

.....

The Strategy or Strategies I have used, Number(s) : .....
If Number "10. / 9. Other Strategy", please specify (in English or in Greek) : .....
.....
.....

*Please circle **one** of the three numbers below to indicate the degree of certainty of your choice of Strategy/Strategies :* *1.* *2.* *3.*

6. What did the results of the experiment indicate?

.....

.....

.....

The Strategy or Strategies I have used, Number(s) : .....
If Number "10. / 9. Other Strategy", please specify (in English or in Greek) : .....
.....
.....

*Please circle **one** of the three numbers below to indicate the degree of certainty of your choice of Strategy/Strategies :* *1.* *2.* *3.*

BEST COPY AVAILABLE

Retrospective Questionnaire

The following questions are to be answered after taking the test.

I. Please read the following statements and **tick** the one/ones that apply to you:

- ❑ I read the whole text first and then began to answer the questions.
- ❑ I read part of the text first and then began to answer the questions.
- ❑ I read all the questions first and then I read the text.
- ❑ I read some of the questions first and then I read the text.
- ❑ First of all I started reading the questions, one at a time and tried to answer them by reading the text to find the answer to each one.
- ❑ I answered the questions in chronological order.
- ❑ I did not answer the questions in chronological order.

(If ticked, please specify your order by giving the numbers of the questions:.....
.....)

II. Please answer the following questions briefly:

1. Which questions did you find difficult to answer and why?

.....
.....
.....

2. Which questions did you find easy to answer and why?

.....
.....
.....

3. Were there any questions that you consider "unfair"? Yes / No.
If Yes, please try to explain why: .....
.....
.....

4. Are there any comprehension questions for which you know the answer but had difficulty answering in English? Yes / No.
If Yes, please answer the question(s) in Greek in the space provided : .....
.....
.....

III. Please think about the following and give your opinion in brief :

1.How do you generally feel about English Language Tests?.....
.....

2. How do you generally feel about reading comprehension tests?.....
.....

3. Do you think that this particular type of exercise (free-response format) tested your reading comprehension of the passage? Yes / No
(If No, why do you think this is so?.....
.....

4. Do you feel that the information required from you after every question (that is, all the extra questions about *Strategies*) interfered with your test-taking process? Yes / No.
(If Yes, in what ways?.....
.....
.....)

Thank you very much for having taken part in this test.

TEST METHOD EFFECT

VERSION B / BI

Preliminary information

This is a test of reading comprehension. Before taking it, please **read all the instructions carefully.**

While taking the test, you will do the following **three** things:

Step 1: Answer the Questions by circling the correct answer.

Step 2: When finishing answering **each** question, decide, based on the list of *Strategies* (given on the following page), which *Strategy* you have used in order to give an answer to this particular question. Beside each question write down the number of the *Strategy* which you used to answer it. If you used more than one *Strategy*, add that too.
If the *Strategy* that you used for a particular question is not on the list, then next to number "13./ 12. Other Strategy " specify what other *Strategy* you used by giving a brief explanation (in English or in Greek).

Step 3: After finishing Step 1 and Step 2 , please rate how certain you are of the *Strategies* you have used to answer the questions. In doing this, you only need to circle one of the three numbers provided at the end of every question. What follows is the description of what every number stands for.

1. = I am certain that I have used the Strategy (-ies) I specified in Step 2.

2. = I am somewhat uncertain that I have used the Strategy (-ies) I specified in Step 2.

3. = I am uncertain that I have used the Strategy (-ies) I specified in Step 2.

Please keep in mind the following : If you make any changes in Step 1, that is, if you change your answer for the comprehension question, remember to change Steps 2 and 3, as these are affected by your previous change as well.

Read the following passage and choose (a), (b), (c) or (d) providing at the same time the information requested in Steps 2 and 3. You can underline or keep notes if you want.

A team of researchers has found that immunizing patients with bee venom instead of with the bees' crushed bodies can better prevent serious and sometimes fatal sting reactions in the more than one million Americans who are hypersensitive to bee stings. The crushed-body treatment has been standard for fifty years, but a report released recently said that it was ineffective. The serum made from the crushed bodies of bees produced more adverse reactions than the injections of the venom did.

The research compared results of the crushed-body treatment with results of immunotherapy that used insect venom and also with results of a placebo. After six to ten weeks of immunization, allergic reactions to stings occurred in seven of twelve patients treated with the placebo, seven of twelve treated with crushed-body extract, and one of eighteen treated with the venom.

1. What is the main topic of the passage?

(a) A new treatment for people allergic to bee stings
(b) A more effective method of preventing bee stings
(c) The use of placebos in treating hypersensitive patients
(d) Bee venom causing fatal reactions in hypersensitive patients

(Refer to Step 2 and 3 of the Preliminary Information Section if you are not sure of how to complete the following part).

The Strategy or Strategies I have used, Number(s) : .....
If Number "13. / 12. Other Strategy", please specify (in English or in Greek) : .....
.....
.....

*Please circle **one** of the three numbers below to indicate the degree of certainty of your choice of Strategy Strategies : 1. 2. 3.*

BEST COPY AVAILABLE

2. According to the researchers, the traditional treatment for bee stings is

(a) widespread
(b) extremely harmful
(c) almost useless
(d) sensitizing

The Strategy or Strategies I have used, Number(s) : .....
If Number "13. / 12. Other Strategy", please specify (in English or in Greek) : .....
.....
.....

*Please circle **one** of the three numbers below to indicate the degree of certainty of your choice of Strategy/Strategies : 1. 2. 3.*

3. The number of patients who took part in the experiment described was

(a) one million
(b) forty-two
(c) twenty-four
(d) eighteen

The Strategy or Strategies I have used, Number(s) : .....
If Number "13. / 12. Other Strategy", please specify (in English or in Greek) : .....
.....
.....

*Please circle **one** of the three numbers below to indicate the degree of certainty of your choice of Strategy Strategies : 1. 2. 3.*

4. The most successful treatment described in the passage was a serum prepared from

(a) the blood of patients who had been stung
(b) poison extracted from bees
(c) crushed bodies of bees
(d) a placebo and a crushed-body extract

The Strategy or Strategies I have used, Number(s) : .....
If Number "13. / 12. Other Strategy", please specify (in English or in Greek) : .....
.....
.....

*Please circle **one** of the three numbers below to indicate the degree of certainty of your choice of Strategy/Strategies : 1. 2. 3.*

BEST COPY AVAILABLE

5. In order to be successful, the treatment referred to in the passage must be administered

(a) by a series of injections given before the patient is exposed
(b) by injection immediately after the patient has been stung
(c) orally for six to ten weeks before the patient is stung
(d) orally immediately after the patient is stung

The Strategy or Strategies I have used, Number(s) : .....
If Number "13. / 12. Other Strategy", please specify (in English or in Greek) : .....
.....
.....

*Please circle **one** of the three numbers below to indicate the degree of certainty of your choice of Strategy/Strategies :* *1.* *2.* *3.*

6. Results of the experiment indicated that

(a) patients treated with venom were stung less frequently
(b) immunotherapy was effective for all patients
(c) immunization took place in seven out of twelve patients
(d) the traditional treatment was as effective as the placebo

The Strategy or Strategies I have used, Number(s) : .....
If Number "13. / 12. Other Strategy", please specify (in English or in Greek) : .....
.....
.....

*Please circle **one** of the three numbers below to indicate the degree of certainty of your choice of Strategy/Strategies :* *1.* *2.* *3.*

BEST COPY AVAILABLE

Retrospective Questionnaire

The following questions are to be answered after taking the test.

I. Please read the following statements and **tick** the one/ones that apply to you:

- ❑ I read the whole text first and then began to answer the questions.
- ❑ I read part of the text first and then began to answer the questions.
- ❑ I read all the questions first and then I read the text.
- ❑ I read some of the questions first and then I read the text.
- ❑ First of all I started reading the questions, one at a time and tried to answer them by reading the text to find the answer to each one.
- ❑ I answered the questions in chronological order.
- ❑ I did not answer the questions in chronological order

(If ticked, please specify your order by giving the numbers of the questions:.....
.....)

II. Please answer the following questions briefly:

1. Which questions did you find difficult to answer and why?
.....
.....
.....

2. Which questions did you find easy to answer and why?
.....
.....
.....

3. Were there any questions that you consider "unfair"? Yes / No.
If Yes, please try to explain why: .....
.....
.....

III. Please think about the following and give your opinion in brief :

1. Do you think that this particular type of exercise (multiple-choice format) tested your reading comprehension of the passage? Yes / No
(If No, why do you think this is so?.....
.....)

2. Do you think that the provision of options to choose from, gave you ideas or alternatives you might have not thought of? Yes / No.

3. Do you think this had a ❑ confusing effect? (please tick your choice)
❑ facilitating

4. Do you feel that the information required from you after every question interfered in your test-taking process? Yes / No.

(If Yes, in what ways?.....

.....

.....

5. Which one of the two formats: ❑ **free-response format**
❑ **multiple-choice format**

do you think tested your reading comprehension of this passage **best**?

* Why do you think this is so? Please give your reasons:.....

.....

.....

6. On which of the two tests do you think you did best?

❑ Free-response version

❑ Multiple-choice version

7. Do you think that seeing the passage twice has helped you understand it better?

❑ Yes

❑ No

8. Do you think that seeing the passage twice has affected your final test score?

❑ Yes

❑ No

9. Do you think that you have become more aware of your *Test-taking Strategies* after taking part in these two tests?

❑ Yes

❑ No

(Could you please explain your choice? : .....

.....

.....)

10. Please add any further comments: .....

.....

Thank you very much for having taken part in this test, too.

VERSION A

Please read the following test-taking Strategies ***carefully*** *and* ***then do the test*** *on the next page. (It was felt convenient and a less intervening factor during the test-taking process if the Strategies were written in the mother tongue of the test takers).*

Checklist of test-taking Strategies

No 1. Guess = I tried to guess the answer without any particular considerations.

No 2. BK knowl. = I used my background knowledge outside the passage.

No 3. Chronolog. = I looked for the answer in chronological order in the passage and on finding an acceptable one, I made a note of it and terminated research.

No 4. The whole = I looked for the answer in the passage and although I found an acceptable one, I did not terminate research but I made a note of this answer as soon as I had finished reading the whole of the passage.

No 5. Locate = After reading the question, I immediately located the area in the passage that the question referred to and then started looking for clues to the answer in that context.

No 6. Match text = I tried to match a word/words/phrase in the question with the same/similar one(s) in the passage.

No 7. Memory = I tried to give an answer based on what I could remember from the passage rather than the passage itself.

No 8. Clues = I received clues from answering another question that helped me answer this one, too.

No 9. Return later = I skipped this question because I could not understand it / could not find an answer to it for the time being and returned to it later.

No 10. Other = I used another Strategy.

VERSION AI

Please read the following test-taking Strategies ***carefully*** *and* ***then do the test*** *on the next page. (It was felt convenient and a less intervening factor during the test-taking process if the Strategies were written in the mother tongue of the test takers).*

Checklist of test-taking Strategies

No 1. Guess = I tried to guess the answer without any particular considerations.

No 2. BK knowl. = I used my background knowledge outside the passage.

No 3. Chronolog. = I looked for the answer in chronological order in the passage and on finding an acceptable one, I made a note of it and terminated research.

No 4. The whole = I looked for the answer in the passage and although I found an acceptable one, I did not terminate research but I made a note of this answer as soon as I had finished reading the whole of the passage.

No 5. Match text = I tried to match a word/words/ phrase in the question with the same/similar one(s) in the passage.

No 6. Memory = I tried to give an answer based on what I could remember from the passage rather than the text itself.

No 7. Clues = I received clues from answering another question that helped me to answer this one, too.

No 8. Return later = I skipped this question because I could not understand it / could not find an answer to it for the time being and returned to it later.

No 9. Other = I used another Strategy.

VERSION B

Please read the following test-taking Strategies ***carefully and then do the test*** *on the next page. (It was felt convenient and a less intervening factor during the test-taking process if the Strategies were written in the mother tongue of the test takers).*

Checklist of test-taking Strategies

No 1. Guess = I tried to guess the answer without any particular considerations.

No 2. BK knowl. = I used my background knowledge outside the passage.

No 3. Chronolog. = I looked for the answer in chronological order in the passage and on finding an acceptable one, I made a note of it and terminated research.

No 4. The whole = I looked for the answer in the passage and although I found an acceptable one, I did not terminate research but I made a note of this answer as soon as I had finished reading the whole of the passage.

No 5. Locate = After reading the question and the alternatives, I immediately located the area in the passage that the question referred to and then started looking for clues to the answer in that context.

No 6. Match text = I tried to match a word/words/phrase in the question and the alternatives with the same/similar one(s) in the passage.

No 7. Memory = I tried to give an answer based on what I could remember from the passage rather than the passage itself.

No 8. Clues = I received clues from answering another question that helped me answer this one, too.

No 9. Return later = I skipped this question because I could not understand it / could not find an answer to it for the time being and returned to it later.

No 10. Match stem = I tried to match a word/words/phrase in the question with the same/similar ones in the alternatives.

No 11. Eliminate = I chose one of the alternatives not because it was thought to be correct, but because the others did not seem reasonable, seemed similar or were not understandable.

No 12. Deduction = I chose one of the alternatives through deductive reasoning.

No 13. Other = I used another Strategy.

VERSION BI

Please read the following test-taking Strategies ***carefully*** *and* ***then do the test*** *on the next page. (It was felt convenient and a less intervening factor during the test-taking process if the Strategies were written in the mother tongue of the test takers).*

Checklist of test-taking Strategies

No 1. Guess = I tried to guess the answer without any particular considerations.

No 2. BK knowl. = I used my background knowledge outside the passage.

No 3. Chronolog. = I looked for the answer in chronological order in the passage and on finding an acceptable one, I made a note of it and terminated research.

No 4. The whole = I looked for the answer in the passage and although I found an acceptable one, I did not terminate research but I made a note of this answer as soon as I had finished reading the whole of the passage.

No 5. Match text = I tried to match a word/words/phrase in the question and the alternatives with the same/similar one(s) in the passage.

No 6. Memory = I tried to give an answer based on what I could remember from the passage rather than the passage itself.

No 7. Clues = I received clues from answering another question that helped me answer this one, too.

No 8. Return later = I skipped this question because I could not understand it / could not find an answer to it for the time being and returned to it later.

No 9. Match stem = I tried to match a word/words/phrase in the question with the same/similar ones in the alternatives.

No 10. Eliminate = I chose one of the alternatives not because it was thought to be correct, but because the others did not seem reasonable, seemed similar or were not understandable.

No 11. Deduction = I chose one of the alternatives through deductive reasoning.

No 12. Other = I used another Strategy.

The following is a compilation of what the students reported when using **Other strategy** for all six questions. Some of the students described this strategy in English and some others in Greek.

The italicised quotations are translations in English done by the present researcher. Each quotation is preceded by the code number of the student. The letter in front of the number indicates the versions of the Checklist s/he was given followed by a number.

The brackets with the indicated strategies following each quotation, were labelled by the present researcher

QUESTION 1

Checklist A (2 occurrences)

A12-B21 : *" I did surface reading of the text, read the question, could not find the answer and I re-read the text more carefully this time".* **(= Other)**

A20-BI20 : " I read the passage one more time and tried to give an answer" **(= Other)**

Checklist AI (4 occurrences)

AI6-BI16 : *" As in Strategy No 7. (Clues), I discovered some clues, but these were mainly related to the frequency with which certain words appeared in the text e.g.* immunising, bee *etc. ".* **(= Locate)**

AI10-B10 : *" I re-read the text so as to make sure that the answer is correct"* **(= Other)**

AI20-B20 : *" I looked for specific clues in the text that helped me to give an answer".* **(= Locate)**

AI23-B23 : " Re-read, picking out the main point ". **(= Other, Locate)**

Checklist B (1 occurrence)

B27-AI27 : " I read the alternative answers; from the beginning I was more or less certain about the answer but kept reading the remaining answers and then I chose (a) with no hesitation". **(= Deduction)**

Checklist BI (1 occurrence)

BI29-A29 : *" I tried to find the phrases of the alternatives in the text so as to see if the overall meaning of the sentences in the text coincides with each one of the alternatives. I rejected the ones that did not coincide and kept the right one".* **(= Match text, Deduction)**

BEST COPY AVAILABLE

QUESTION 2

Checklist A (no occurrences)

Checklist AI (2 occurrences)

AI23-B23 : " I searched for a logical answer. After having read the question, returned to the appropriate part of the passage and 'fished out' the answer " **(= Locate)**

AI27-B27 : " I returned to the text and found the answer ". **(= Other)**

Checklist B (no occurrences)

Checklist BI (1 occurrence)

BI20-A20 : " I answered by chance because I did not remember what (d) meant ". **(= Guess)**

QUESTION 3

Checklist A (no occurrences)

Checklist AI (5 occurrences)

AI10-B10 : *" I remembered the text well enough so as to go back to the part I was interested in ".* **(= Locate)**

AI11-B11 : *" The answer is calculated practically from the 2nd paragraph ".* **(= Locate)**

AI15-B15 : " I counted the number of the patients ". **(= Locate)**

AI16-BI16 : " I counted the patients as a whole ". **(= Locate)**

AI23-B23 : " Searched for the logical answer. After having read the question, returned to the appropriate part of the passage and 'fished out' the answer". **(= Locate)**

Checklist B (no occurrences)

Checklist BI (6 occurrences)

BI2-A2 : *" I immediately looked for the answer in the part of the text where the number of the patients was found ".* **(= Locate)**

BI6-AI6 : " I re-read only the part of the text in which I remembered the statistics to be, then I did the counting and tried to find the answer that corresponded to my estimation ". **(= Locate)**

BI12-AI21 : (gave no explanation)

BI15-A15 : " I looked at the text and counted the patients that took part in the experiment ". (= **Locate**)

BI21-A21 : *" I immediately went to the part of the text where I knew that the answer was ".* (= **Locate**)

BI29-A29 : *" I immediately went to the 2nd paragraph and counted the patients ".* (= **Locate**)

QUESTION 4

Checklist A (2 occurrences)

A21-BI21 : " I remembered the answer and went back to the passage just to make sure it was actually the bee venom ". (= **Other**)

A27-B28 : *" I read the question and went back to the text more than one times trying to find the same words / expressions ".* (= **Other**)

Checklist AI (2 occurrences)

AI23-B23 : " Searched for the logical answer. After having read the question, returned to the appropriate part of the passage and 'fished out' the answer ". (= **Locate**)

AI22-B22 : (gave no explanation)

Checklist B (1 occurrence)

B27-AI27 : " I read the alternative answers, from the beginning I was more or less certain about the answer but kept reading the remaining answers and then chose (b) with no hesitation ". (= **Deduction**)

Checklist BI (2 occurrences)

BI6-AI6 : " I read the text in order to make certain ". (= **Other**)

BI12-AI21: (gave no explanation)

QUESTION 5

Checklist A (1 occurrence)

A8-BI8 : *" I skipped this question because I could not find an answer in the text ".* (= **Other**)

Checklist AI (two occurrences)

AI23-B23 : " Searched for the logical answer. After having read the question returned to the appropriate part of the passage and 'fished out' the answer ". **(= Locate)**

AI4-B4 : *" I used clues without being absolutely sure if this is the appropriate answer ".* **(= Locate)**

Checklist B (1 occurrence)

B4-AI4 : (gave no explanation)

Checklist BI (2 occurrences)

BI20-A20 : " I skipped it because I think this exact piece of information does not exist in the text ". **(= Other)**

BI29-A29 : *" I went to the corresponding sentence in the text and tried to compare its meaning with all the alternatives and by rejecting the wrong ones I arrived at the correct one ".* **(= Locate, Match text, Deduction)**

QUESTION 6

Checklist A (no occurrences)

Checklist AI (2 occurrences)

AI20-B20 : *" I immediately re-read the part where I remembered the answer to be ".* **(= Locate)**

AI23-B23 : " Searched for the logical answer. After having read the question returned to the appropriate part of the passage and 'fished out' the answer ". **(= Locate)**

Checklist B (2 occurrences)

B23-AI23 : " Process of elimination ". **(= Eliminate)**

B27-AI27 : " I sort of guessed the answer ". **(= Guess)**

Checklist BI (two occurrences).

BI2-A2 : *" I re-read only the last sentence because I remembered where exactly the result of the research was ".* **(= Locate)**

BI21-A21 : *" I answered the question because I could remember the part of the text it was included but I went back to it for verification ".* **(= Locate, Other)**

The following is what Subjects A and C reported during the introspective interviews. These were conducted in Greek. Whenever English was used, it is highlighted in the protocols below. Only the parts of the protocols related to how the Subjects gave the answers to the six questions is included below.

QUESTION 1

Free-response format:

Subject A: " ..... I realised that this was in the 1st paragraph where I returned to; this was in a sentence, so I retained this in my memory and I gave an answer ..... I remembered that this was in the 1st paragraph ..... I went to the 1st paragraph because I was not sure about the wording of my answer **(= Locate)** ..... I read the second paragraph so as to be more sure and then I ended up in the 1st paragraph ....." **(= The whole)**.

Subject C: " I looked at the 1st paragraph, the 1st sentence so as to see how exactly it is phrased..... I could have answered with my own words but because I wanted to confirm it and because I also wanted to figure out the way to express it, I looked at the 1st sentence ....." **(= Locate)**

Multiple-choice format :

Subject A: " ..... the first answer seems closer to the question ..... (b) is not, because the text doesn't say anything about *preventing bee stings*, (c) was a *topic* but not *the main topic* because this is the *venom* ..... and (d) that contains the *venom* has nothing to do with *causing fatal reactions*. So it is (a) ". **(= Deduction)**

Subject B: " ..... I am going to choose (a); (b) is excluded because it is about *preventing bee stings* ..... well (c) is not ..... our main topic, the placebos and ..... the *main topic* is not (d) , it is the new treatment ". **(= Deduction)**

QUESTION 2

Free-response format:

Subject A: " This question is related to the answer in the first question **(= Clues)** ..... I am now going to the text to find the answer ..... I remember it is in the 1st paragraph and more particularly in the 5th line where it says that *the crushed-body*

treatment has been standard for fifty years, but a report released recently said it was ineffective. (= **Locate**) .....but I am reading the 1st paragraph again ..... I read the 2nd paragraph quickly ..... but this has nothing to do with it ..... it is about another thing (= **The whole**) ..... I am going back to the text to find out how to compose my own sentence, to see the terminology used in line 5 ..... the word *report* ..... I looked at the text to find out when this *report* took place which is *recent* ". (= **Other**)

Subject C: " ..... I looked here : *The crushed-body treatment has been standard for fifty years, but a report released recently said it was ineffective*. I didn't look anywhere else in the text. I referred to this particular part because I knew that the answer was there ". (= **Locate**)

Multiple-choice format:

Subject A: " ..... but the answer must be (c) because the text says it is *ineffective* (= **Locate**) ..... So *ineffective* is not *widespread* nor *extremely harmful* nor *sensitising*. It is *almost useless* ". (= **Match text**)

Subject C: " ..... I am going back to the text ..... I am starting reading from the beginning, from the 1st paragraph and ..... now I've stopped at *ineffective* . (= **Chronolog.**) ..... Now I am thinking if (a) or (c) is the correct answer but *almost useless* ..... yes, I don't think it is this one, probably *widespread* from what I have read ..... according to the text : *The crushed-body treatment has been standard for fifty years* (= **Match text**) ..... (b) is contrary to the whole meaning of the text, so I reject this ..... and *sensitising* I also reject this one because it is irrelevant to the overall meaning of the text " . (= **Deduction**)

QUESTION 3

Free-response format:

Subject A: " I am now going to the 2nd paragraph ..... but I am just looking at the top as well but I am reading the 2nd paragraph more carefully ..... The 2nd paragraph is about the patients .....yes. The last sentence is about the patients, what happened to them and how they were used that is they were used in dozens (= **Locate**) ..... and I am going back to the 1st paragraph ..... It doesn't say anything about how the patients were used (= **The whole**) so ... 12+12=24+18=42 "

Subject C: " I returned to the text, I counted how many the people were and ..... I went to the last paragraph. I didn't go to the 1st paragraph at all " . (= **Locate**)

Multiple-choice format:

Subject A: " I am now reading the last paragraph : 12+12=28+18=42. It is (b) because of arithmetic " . (= **Locate**)

Subject C: " I reject *one million* because I remember the number is not that big but the number of the people who took part in the experiment is less than that (= **Deduction**) so I am now looking at the text to find the exact number ..... I am looking at the 2nd paragraph, last line, where the relevant information is from what I can remember .....12 patients, 12, 24+18=42 yes 42 " . (= **Locate**)

QUESTION 4

Free-response format:

Subject A: " Now I am reading the 4th question. At this very moment I am thinking of the 3rd question and it is really awful because I was thinking about it while I was reading the 4th just to see if I could relate it in any way to the 4th, to see if I can find something because ..... yes, it can be related because from those 42 I have to look back in the paragraph in which they are mentioned per dozens and those 18 afterwards to find out how each one of the treatments was used and what their results were ..... (= **Clues**) So I am going to read this again to find out the exact findings (= **Locate**) ..... I am constantly looking, up and down the 2nd paragraph and the 4th question ..... Now I am looking at the 1st paragraph a little bit, because it is about how they made the serum and what they gave to them and I want to find out if there are any clues there ..... Now I am going to the 2nd paragraph ..... I am going to the 1st paragraph, 1st sentence to find out about the use of *venom* and I am finally concluding that this is the answer I have to give " . (= **The whole**)

Subject C: " I looked at the 1st paragraph and I found out about the *bee-venom*. I didn't read the whole of the 1st paragraph. I only read the 1st sentence ..... (= **Locate**). I returned to the text so as to make sure although I could have answered the question without doing this " . (= **Other**)

Multiple-choice format:

Subject A: " ..... (a) is a little bit macabre. It can't be this one (laughs) ..... These ones are a little bit tricky. *The blood* is definitely not ..... *poison?* ..... *extracted from bees?* ..... (c) *crushed bodies of bees* is not because as we said before the *crushed bodies* were ineffective (=**Deduction**) ..... I am looking again at the 1st paragraph. I am reading the 1st and the 2nd paragraph (= **The whole**) ..... It is (b) because .....

(a) is not because we don't use the *blood of the patients*, we use the *bees* in general. The *crushed-bodies*, as we said, is ineffective; it was the previous treatment and (d) which is the *placebo* and the *crushed-body extract*, we have already rejected the *crushed-body* in the previous one, which makes this answer almost half, so it is *poison extracted from bees*, that is the *venom* . That is we eliminated one after the other and we kept the one that is left ". **(= Deduction)**

Subject C: " (a) is completely irrelevant. It is rejected by using common sense that is I am rejecting it according to the text. I keep (b). It must be the right one; (c) ..... according to the text it is the traditional treatment which was not sufficient, was not successful and the last one is rejected also according to the text ;**(= Deduction)** according to what I remember " . **(= Memory)**

QUESTION 5

Free-response format:

Subject A: " I am reading number 5 now ..... and I have already read number 6 ..... because there is something in number 5 that I do not understand **(= Return later)** ..... I am reading number 5 again ..... I am going to the text from the beginning to find out ..... I read the 1st paragraph **(= Chronolog.)** which says something about 6. Number 6 has something to do with 5 and 3 **(= Clues)** . I am going back to 6. The answer for number 6 must be at the beginning of the 1st sentence again, as for *the main topic* **(= Clues)** ..... I am reading the 1st paragraph from the 5th line onwards ..... I am going to 5 ..... ".

Subject C: " I looked at different parts of the text. I am not sure if I found the answer in a specific part. That is I didn't go directly to the part it was mentioned. I looked for the answer in the 2nd paragraph, I realised that it was nowhere there to be found, and then I went to the 1st paragraph and finally I found it in *serum made from the crushed bodies of bees* ..... er and *injections of the venom*" . **(=The whole)**

Multiple-choice format:

Subject A: " Two answers are about after the patient is stung and the other two are about before. I am going back to the text again ..... Now I am reading the 2nd paragraph : *After six to ten weeks of immunisation* **(= Locate)** ..... but ..... it doesn't say anything about when it took place, when these things happened, if it was before or after the patients were stung ..... (c) is the only answer that contains *six to ten weeks* which is exactly the same thing with what it says here ..... *After six to ten*

weeks of immunisation (= **Match text**) so it is before; it is so because it says *before the patient is stung* because after *six to ten weeks* of immunisation they took them to the bees and *allergic reactions occurred accordingly*. So here I rejected the answers that had to do with whether the treatment was after they were stung ..... which was not correct from the beginning ..... and I ended up with (c) because it has *six to ten weeks* and of course it says *before* and most probable it is *orally* because the text doesn't say anything about *injections* which up to this point they were *ineffective* ..... but it is not clear at all ..... (= **Elimination**). One has to go through a particular mental process, that is to say: it is not this one nor that one, so it must be the third one" . (= **Deduction**)

Subject C: " I think it is nowhere. The text doesn't contain such details but I am going to have a look again ..... (starts reading from the beginning of the text) ..... I don't think it is here ..... (starts reading the 2nd paragraph) *six to ten weeks of immunisation* (= **The whole**)..... Let's see. I think it is (c) according to *six to ten weeks of immunisation* (= **Match text**) ..... So it is (c); (a) doesn't mention anything about time. Here we have six to ten weeks . It is too general. I wouldn't go for that; (b) ..... no, it isn't according to what I have read , to what the text is about. (=**Deduction**) ..... I am confused ..... *Orally* is through the mouth. In the 1st paragraph it says that *injections of the venom did* . (= **Match text**) So it is an injection. So *orally* doesn't fit here. So it is (a). I was confused ..... (a) " . (= **Deduction**)

QUESTION 6

Free-response format:

Subject A: " I believe that the answer is again in the 1st sentence (= **Locate**) where I had found the *main topic*. (= **Clues**) I am reading little by little the 1st sentence of the 1st paragraph to find out about the wording of the results ..... (starts writing) ..... I am reading the 1st sentence to have a look at the wording (= **Other**) ..... (while writing :) There is one piece that I remember it so well that I do not have to go back to the text because I remember it very well ; a piece, small enough which is in the 1st sentence of the 1st paragraph " .(= **Memory**)

Subject C: " I thought of the answer and I wrote it. I didn't have to go back to the text. This time I was sure that the *results* were clear enough in the text and I had them in my mind clearly enough ..... " . (= **Memory**)

Multiple -choice format:

Subject A: " *The results* ..... The first answer is not correct one way or another because it doesn't have to do with how you are stung ..... Here they are a little bit tricky; (d) ? It is true that the *traditional treatment was as effective as the placebo*. It had exactly the same results. But this is not the only *result of the experiment*. Most probable (= **Deduction**). Neither do I have to go back to the text and check the rest (= **Memory**). I am going to (c) ..... and I going to the 2nd paragraph to find out the exact number (= **Match text**). It is not true because *immunisation was effective for all patients*. Nor is it (a). That was excluded from the beginning ..... So it is (d). Although the *results* were ..... there was a better result for those who had taken the *venom* . That was the *result* of the *experiment*. As a secondary result one could say that .....yes, *the traditional treatment was as effective as the placebo* because it had exactly the same results. Out of 12 patients 7 had exactly the same symptoms. In both treatments". (= **Elimination**)

Subject C: " I am looking at the text now, at the last sentence which is about the results of the research ..... *allergic reactions occurred in seven of twelve patients treated with the placebo, seven of twelve treated with the crushed-body extract* (= **Locate**) so it is ..... (d) *the traditional treatment was as effective as the placebo*; (a) is rejected because it says *were stung less frequently*. That is irrelevant; (b) is rejected because there is no such information in the text and (c) doesn't fit ". (= **Deduction**)

Facility value and their difficulty level as perceived by the students.

Table 1. Free-response items

QUESTIONS	FV	FE (a)	FD (b)
1	.59	31	8
2	.89	40	2
3	.98	40	5
4	.52	40	4
5	.31	5	41
6	.78	37	4

Table 2. Multiple-choice items

QUESTIONS	FV	FE (a)	FD (b)
1	.82	37	0
2	.70	26	14
3	.94	42	3
4	.87	40	4
5	.59	19	22
6	.63	23	15

(a) FE = frequency of students who claimed that the question was easy.
(b) FD = frequency of students who claimed that the question was difficult.

Facility value and their difficulty level as perceived by the students.

Free-response items

QUESTIONS	FV	FE (a)	FD (b)
1	.59	31	8
2	.89	40	2
3	.98	40	5
4	.52	40	4
5	.31	5	41
6	.78	37	4

2. Multiple-choice items

QUESTIONS	FV	FE (a)	FD (b)
1	.82	37	0
2	.70	26	14
3	.94	42	3
4	.87	40	4
5	.59	19	22
6	.63	23	15

(a) FE = frequency of students who claimed that the question was easy.
(b) FD = frequency of students who claimed that the question was difficult.

Overall frequency of Strategy use of Free-response and Multiple-choice format for all six questions *

	QUESTION 1				QUESTION 2				QUESTION 3				QUESTION 4				QUESTION 5				QUESTION			
Strategies	A	AI	B	BI	A	AI	B	BI	A	AI	B	BI	A	AI	B	BI	A	AI	B	BI	A	AI	B	BI
Guess	4	3	5	2	1	1	2	3	0	3	1	3	3	1	2	3	2	0	1	0	2	1	3	0
BK knowl.	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	4	1	1	0	1	1
Chronolog.	1	5	1	1	7	14	2	6	2	10	2	12	1	4	2	5	2	6	1	6	1	2	1	2
The whole	5	7	5	4	1	5	2	3	1	5	2	3	5	5	2	3	7	4	6	6	1	4	3	6
Locate	14		6		22		11		25		18		13		16		11		10		14		12	
Match text	2	1	0	3	2	3	2	4	0	2	2	3	0	2	1	3	2	1	1	4	1	1	0	3
Memory	6	11	13	11	1	4	4	4	0	4	3	1	8	15	5	10	3	3	1	3	11	17	5	5
Clues	1	0	0	0	0	0	0	0	1	0	0	0	2	1	0	0	0	0	0	0	2	3	0	0
Return later	1	0	0	0	0	1	6	2	1	0	0	0	0	0	0	0	6	10	1	1	0	0	0	0
Other	2	4	1	1	0	2	0	1	0	5	0	6	2	2	1	2	1	2	1	2	0	2	2	2
Match stem			0	1			1	0			0	1			0	1			0	2			0	0
Eliminate			0	0			2	2			1	0			1	0			1	0			0	4
Deduction			4	9			2	6			0	2			0	5			3	7			8	13
Total	**36**	**31**	**35**	**32**	**34**	**30**	**34**	**31**	**30**	**30**	**29**	**31**	**34**	**30**	**30**	**32**	**34**	**28**	**30**	**32**	**33**	**30**	**35**	**36**

* The boxes in which there is no mention of strategy use, indicate that the strategy / strategies for this Checklist have been deleted.

Table 4. 2. 1. Overall frequency of Strategy use of Free-response and Multiple-choice format for all six questions *

	QUESTION 1				QUESTION 2				QUESTION 3				QUESTION 4				QUESTION 5				QUESTION			
Strategies	**A**	**AI**	**B**	**BI**	**A**	**AI**	**B**	**BI**	**A**	**AI**	**B**	**BI**	**A**	**AI**	**B**	**BI**	**A**	**AI**	**B**	**BI**	**A**	**AI**	**B**	**BI**
Guess	4	3	5	2	1	1	2	3	0	3	1	3	3	1	2	3	2	0	1	0	2	1	3	0
BK knowl.	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	4	1	1	0	1	1
Chronolog.	1	5	1	1	7	14	2	6	2	10	2	12	1	4	2	5	2	6	1	6	1	2	1	2
The whole	5	7	5	4	1	5	2	3	1	5	2	3	5	5	2	3	7	4	6	6	1	4	3	6
Locate	14		6		22		11		25		18		13		16		11		10		14		12	
Match text	2	1	0	3	2	3	2	4	0	2	2	3	0	2	1	3	2	1	1	4	1	1	0	3
Memory	6	11	13	11	1	4	4	4	0	4	3	1	8	15	5	10	3	3	1	3	11	17	5	5
Clues	1	0	0	0	0	0	0	0	1	0	0	0	2	1	0	0	0	0	0	0	2	3	0	0
Return later	1	0	0	0	0	1	6	2	1	0	0	0	0	0	0	0	6	10	1	1	0	0	0	0
Other	2	4	1	1	0	2	0	1	0	5	0	6	2	2	1	2	1	2	1	2	0	2	2	2
Match stem			0	1			1	0			0	1			0	1			0	2			0	0
Eliminate			0	0			2	2			1	0			1	0			1	0			0	4
Deduction			4	9			2	6			0	2			0	5			3	7			8	13
Total	**36**	**31**	**35**	**32**	**34**	**30**	**34**	**31**	**30**	**30**	**29**	**31**	**34**	**30**	**30**	**32**	**34**	**28**	**30**	**32**	**33**	**30**	**35**	**36**

* The boxes in which there is no mention of strategy use, indicate that the strategy / strategies for this Checklist have been deleted